

## When the fighting is over...

**The conflict in Iraq has divided world opinion, driving deep wedges even between longstanding allies. In its wake, rebuilding international collaboration will be vital — and in that task, scientists should take a lead.**

**W**ill Saunders, a British astronomer at the Anglo-Australian Observatory in Sydney, left the world in no doubt as to his feelings about the current situation in Iraq. He is one of two protesters arrested for daubing “No war” in huge red letters on the roof of Sydney Opera House (see page 366). In January, 41 Nobel laureates chose a more conventional means of making a similar point, declaring that an attack on Iraq without widespread international support “would undermine, not protect, US security and standing in the world”.

Within the broad church of the scientific community, however, there will be those who believe, with equal sincerity, that war is justified to rid the world of a regime that has repeatedly attacked its neighbours and used chemical weapons against its own people. Public statements by pro-war scientists are difficult to find, but some researchers will privately feel that the real villains of the piece are those nations that refused to lend international legitimacy to military intervention. Other scientists will be watching the hostilities unfold in unprecedented detail on their televisions, unsure what to make of it all.

Given this spectrum of opinion, bodies that represent researchers have mostly been silent, realizing that there is no single ‘scientific’ perspective. Nonetheless, the scientific community should recognize that it has some particular responsibilities, and a huge stake in rebuilding an international consensus. Scientists, after all, helped to develop both the high-tech armaments being used to pursue the assault on Iraq, and the weapons of mass destruction whose alleged possession by Saddam Hussein paved the road to conflict in the first place. Scientists will also play a leading role in verifying claims by the

combatants about both types of weapon (see pages 362–363).

More fundamentally, science owes its rapid advancement to the free exchange of ideas and personnel between labs across the globe. Security restrictions introduced by the US administration as part of its ‘war on terror’ are already impeding this exchange (see *Nature* **422**, 96–97; 2003). If the situation is not now to deteriorate further, scientists must speak out, both to protect progress in their own disciplines and to defend a plethora of international agreements thrashed out under the auspices of the United Nations. The Kyoto Protocol on climate change, the Convention on Biological Diversity, the Chemical Weapons Convention, the Nuclear Non-Proliferation Treaty and others all owe their existence, at least in part, to international scientific collaboration.

The world has much to lose if the current conflict leaves this legal framework, like central Baghdad, in smouldering ruins. Even before the Iraqi crisis came to a head, President George W. Bush’s administration had shown itself to be no lover of international agreements. Now, having split so acrimoniously from many of its traditional allies, there is a danger that the world’s only superpower may decide that there is little to be gained by working with the United Nations. And that could, in turn, give rise to a highly dangerous situation in which nations and religious factions jostle freely in pursuit of their own interests.

Instead, we must hope that an international consensus can be rebuilt, if falteringly at first. Scientific organizations have not had much to contribute to the debate running up to this war. But in its aftermath, they should speak up loud and clear to press for internationalism in a world that could otherwise veer towards factionalism and further conflict. ■

## In praise of good mentors

**Scientific mentoring gets less recognition than it deserves, not least for its potential in helping researcher minority groups.**

**T**he fog of war last week obscured something positive that President George W. Bush’s administration was trying to promote — the recognition of outstanding scientific mentors. Ten individuals and six institutions were honoured last week in the 2002 Presidential Awards for Excellence in Mathematics, Science and Engineering Mentoring at a White House ceremony.

Mentoring as a component of one’s scientific career is one of those things that most would agree are positive — at least in the abstract. But mentoring, in the concrete, is rarely recognized beyond the gratitude of the young scientists towards a senior figure who (unless they were unlucky in their choice of laboratory) helped them to publish their first paper or land their first faculty position.

The ten awardees have made a welcome escape from this vacuum of recognition. Perhaps more importantly, their efforts show that seemingly intractable problems in the make-up of the scientific and technological workforce can be addressed. For example, Robert Gray, one of the awardees, guided 11 female students towards PhDs in his electrical engineering programme at Stanford University over 16 years (for information on more awardees, see [www.nsf.gov](http://www.nsf.gov)). In the same time period, 16 men in the programme received the degree. Gray’s results — and his recognition for achieving them — are

notable because the rate of female representation among his doctoral graduates is almost four times the norm in engineering. According to the US National Science Foundation’s Science and Engineering Indicators 2002, engineering has the lowest representation of women — about 10% in all.

Gray’s success on a small scale begs the question of what more successful mentoring could do on a larger scale — particularly for women and minorities. Despite progress, both groups are still highly underrepresented in science, according to recent reports. In the United States, for example, blacks, hispanics and native Americans together make up only 7% of the scientific academic workforce, even though they represent 24% of the population, according to the US National Science Foundation’s indicators. And men far outnumber women in the physical sciences, according to the International Union of Pure and Applied Physics, which last year published a report on gender disparity.

Recognizing mentors through awards such as those presented last week is one way to encourage scientists to give guidance to their younger colleagues — and also points young scientists towards outstanding labs. Similar recognition by scientific societies and other governments could thrust mentoring further into the spotlight — and help to solve some social problems in the process. ■





**On target?**  
Experts seek to assess  
precision of US  
military offensive  
**p362**



**Drowned out**  
World water summit  
overshadowed by  
conflict in Iraq  
**p364**



**Moderate view**  
Sherwood Boehlert  
edges science into  
Congress's agenda  
**p365**



**High note**  
Astronomer faces  
deportation over  
graffiti  
**p366**

# Academics fume as university refuses to reject tobacco dollars

**Rex Dalton, San Diego**

Academics at the University of California are on a collision course with their administrators this week over a proposal to close the door on the acceptance of research funds from tobacco corporations.

Faculty groups at two of the university system's research campuses have voted for such a ban — partly so they can obtain grants from anti-tobacco funding agencies that only donate to institutions with such a rule in place.

But the votes, at the San Francisco (UCSF) and San Diego (UCSD) campuses, are being resisted by university administrators, who do not want to set a precedent for refusing support from other funding sources, such as pharmaceutical companies.

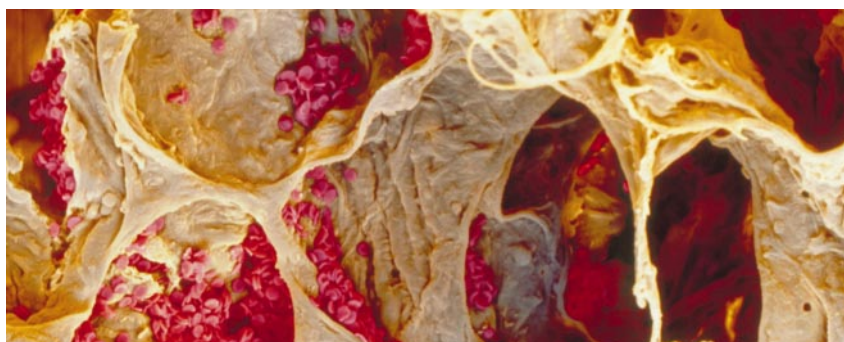
The bans would infringe academic freedom, argued UCSF anaesthesiologist Jeanine Wiener-Kronish in *UCSF Today* last month. "There is a danger when limits on research funding are proposed," she says, "as various groups have political agendas."

In a January vote on UCSF's health-science campus, however, researchers voted 52% to 48% in favour of a ban, with about one-third of the 1,800 eligible faculty voting. Stan Glantz, a UCSF cardiovascular researcher and ban supporter, says: "It is not appropriate to take money from an industry that kills 5 million people worldwide and constantly lies."

Lawrence Coleman, a physicist and vice-provost for research at the University of California, says he is concerned that accepting grants from agencies that make restrictive stipulations could infringe academic freedom.

The issue has been brought to a head by the American Legacy Foundation — a non-profit organization set up in Washington DC as part of a multi-billion-dollar settlement in 1998 between US state governments and the tobacco industry, to fund tobacco research and education. The foundation donates about \$25 million a year to researchers, but only to institutions that expressly ban tobacco money. This policy, it says, is a response to the tobacco companies' record of manipulating research findings and inaccurately claiming credit for them.

About 15 US universities have debated the Legacy Foundation's funding restrictions so far, foundation officials say. Ten of them



Many researchers say tobacco firms shouldn't sponsor research on topics such as lung cancer (pictured).

have agreed on a policy that is acceptable to the foundation, and the remainder have declined to accept the organization's grants.

Researchers at the UCSD Cancer Center also voted this month to disallow tobacco research money, but university officials have not yet acted on it. If they fail to do so by next month, David Burns, a physician who studies tobacco control at the centre, is set to lose a \$500,000 grant from the Legacy Foundation. Like UCSF's researchers, staff at the UCSD Cancer Center say that they do not actually receive any funding from tobacco firms.

The impasse has prompted the University of California's systemwide Academic Senate, which represents top academics at all nine of

the university's campuses, to undertake to formulate a single, consistent policy. But this could take a year — during which time no University of California researcher will be eligible for a Legacy Foundation grant. Gayle Binion, a political scientist and chair of the senate, says that such an impasse would be "irrational", and hopes a compromise can be reached.

The Legacy Foundation is not the first to force universities to choose between its own funds and tobacco money. Two years ago, Britain's top cancer charities withheld funds from the University of Nottingham after it accepted £3.8 million (US\$6 million) from British American Tobacco to set up a corporate ethics centre (see *Nature Med.* 7, 5; 2001). ■

## Europe told to unite on spy satellites

**Quirin Schiermeier, Munich**

**The European Union (EU) needs a joint space-based military observation system to allow it to act for itself in global crises, says the union's research commissioner.**

Speaking on 19 March at a conference on European security in Brussels, Philippe Busquin called for a research programme to aid in the development of such a system.

Satellites for military reconnaissance and intelligence gathering are being developed by at least four EU member states — but each is based on independent technologies.

"Security must be a key element of a European space policy," Busquin said.

"There should be no reason why Europe cannot develop the space assets that are fundamental to any credible security policy."

But according to some estimates, a common EU satellite-monitoring system would cost about €8 billion (US\$8.5 billion).

In January, the Greek EU presidency proposed a European agency similar to the US Defense Advanced Research Projects Agency (DARPA), which supports defence projects in many fields (see *Nature* 421, 465; 2003). The European Commission will this year ask representatives from member states, industry and research organizations to draft a defence-research agenda for the EU. ■

# Analysts seek proof of military precision

Geoff Brumfiel, Washington

Amid all the questions surrounding the second Gulf War, one thing is reasonably certain: great claims will be made for the new generation of guided weapons systems that the Pentagon is deploying in Iraq. For a group of sceptical physicists and defence experts, that represents a special challenge.

At outfits such as the Security Studies Program at the Massachusetts Institute of Technology (MIT), these experts are gearing up to second-guess the Pentagon's assessment of its systems. It's a painstaking task, given the paucity of information that will fall into the public domain. But it's also one of considerable political significance. The US and British governments have stated that they want to defeat Saddam Hussein while keeping civilian casualties to a minimum, and without destroying Iraq's economic infrastructure. Meeting those goals will depend crucially on the military's precision munitions working as advertised to hit accurately identified targets.

Defence analysts will also be looking at the performance of a new generation of Patriot missile interceptors, which are close cousins of the Pentagon's ballistic-missile defence system. After the last Gulf War, it took an MIT team led by physicists George Lewis and Theodore Postol two years to debunk the Pentagon's claim that the older Patriot missile system successfully intercepted most of Iraq's Scud missiles. This time, the independent and unwanted assessors are expecting more data from the larger number of reporters and film crews that the US-led coalition is letting in to cover the war. But they also



Shock and awe: independent verification of the accuracy of the US bombardment may be hard to obtain.

fear that a clampdown on other public data could make their task more difficult. "I'm going to be looking at whatever I can get data on," says Lewis.

With Saddam Hussein's ruling party well ensconced in central Baghdad, the coalition's goal depends on 'smart' weapons finding their targets with pinpoint accuracy. Smart bombs generally fall into two categories, guided either by lasers or satellite. From a technical standpoint, the ability of these bombs to strike close to a prescribed target is reasonably well established, according to William Arkin, a

Washington-based defence analyst.

What is far less clear is the quality of intelligence given to the weapons before they are fired, says Barry Posen, a professor of political science at MIT. During the 1999 conflict in Kosovo, the Pentagon claimed that precision munitions destroyed hundreds of Serbian tanks and armoured vehicles. But under closer scrutiny, it became clear that many of the weapons had hit decoy targets. "I think the verified claims in the end were no more than a few dozen," says Posen.

Press reports will be about the only way to

## Inspectors face uphill struggle in search for banned weapons in Iraq

Natasha McDowell

A second invasion of Iraq may soon be under way — this time by military teams of technical experts trained to identify and destroy weapons of mass destruction (WMD).

According to the US Department of Defense, 'assessment teams' will examine a prioritized list of suspected WMD sites. Where they find suggestions of suspicious activity, 'mobile exploitation teams' will be called in to take samples and conduct further analyses in sophisticated mobile laboratories. Finally, 'disablement teams' will be deployed to destroy any biological or chemical weapons that are found.

The Pentagon declines to comment in detail on the technology at the teams' disposal. But the assessment teams are likely to be equipped with portable devices that can rapidly scan for the presence of banned weapons. These include hand-held chemical-agent monitors, which ionize

air samples and then examine the mobility of the clusters of ions that are formed. Another portable device can analyse chemicals inside sealed containers by recording the  $\gamma$ -ray spectrum given off when the contents are bombarded with neutrons.

But experts warn that portable devices — which may also be carried by frontline troops — can give false positive results. This happened several times during the first Gulf War, says Kenneth Boutin, a researcher at VERTIC, a non-governmental organization in London interested in the verification of arms-control agreements. As *Nature* went to press, the dubious reliability of initial reports had already been underlined by a claim on 23 March — later dismissed as "premature" by Pentagon sources — that US troops had found a chemical-weapons facility near Najaf, some 160 kilometres south of Baghdad.

The Pentagon says that its teams may seek

help from members of the United Nations Monitoring, Verification and Inspection Commission (UNMOVIC), the inspection body that until recently was operating in Iraq. "We envisage, at the appropriate time, augmenting our teams in theatre with former UN inspectors," a Pentagon spokesman told *Nature*. According to some reports, several UNMOVIC inspectors have already been approached.

But given that the political justification for the war depends heavily on the assertion that Iraq has continued to possess WMDs in defiance of UN resolutions, the international community may want to see verification of any positive results by laboratories entirely independent of the US and British governments. "Independent verification is vital," says Julian Perry Robinson of the University of Sussex in Brighton, director of the Harvard Sussex Program, which aims to inform public policy on chemical and biological warfare. ■



assess the military's latest intelligence-gathering technologies, says Philip Coyle, a senior adviser for the Center for Defense Information, who spent seven years as director of the Pentagon's Operational Test and Evaluation Directorate. By scrutinizing media coverage and the defence department's public statements, Coyle hopes to study the success of technologies aimed at telling smart bombs where to strike, such as the new computerized command and control systems, and unmanned aerial vehicles.

Postol and Lewis, meanwhile, will be seeking clues about the performance of the Patriot missile interceptor. After the first Gulf War, the MIT team conducted a frame-by-frame analysis of 33 Patriot intercept attempts to determine the altitude, speed and outcome of each engagement. "We were able to get an enormous amount of information from the footage," says Postol. Their analysis showed that at least a third of the Patriots fired during the first Gulf War failed to destroy a Scud missile — demolishing the army's initial claims of a 96% success rate for the system.

This time the MIT researchers plan to videotape hours of news coverage to track the latest generation of Patriot missiles, known as the Patriot Advanced Capability-3 (PAC-3). Unlike the original Patriots, which used explosives to blow up incoming missiles, the PAC-3 destroys warheads by the sheer force of its collision with them. This 'hit-to-kill' technology is also the basis of the first-generation US ballistic-missile defence system being built in Alaska, heightening interest in the system's performance. Already, the army's central command says that four Iraqi missiles have been intercepted by the PAC-3 system.

Some analysts complain that, even as the Pentagon opens up the battlefield to the press — assigning some 500 reporters to US military units — it is shutting down other data sources. Earlier this year, for example, the US administration began classifying missile-defence test data (see *Nature* **417**, 777; 2002). John Pike, head of Washington-based Globalsecurity.org, says that he used to use commercial satellite images to analyse military activity, but when US troops entered Afghanistan in 2001, the United States bought all of the images so that no one could access them. It is unclear whether the administration will follow a similar course during the present conflict with Iraq, he says.

"The data are getting worse and worse," says Arkin. Nonetheless, Lewis says that his team will continue to scour hours of news coverage to find hints about how weapons are performing. "It's more important than ever to do this because this administration has become so secretive," he says. ■



The US may face accusations of hypocrisy if it deploys riot-control agents during street fighting.

## Critics slam US over plans to use riot-control chemicals in the Gulf

Jonathan Knight, San Francisco

As thousands of US troops poured into Iraq last week, an argument was breaking out back home about their possible readiness to use tear-gas or even chemical calming agents during the conflict.

Some military experts argue that riot-control agents could help to reduce enemy and civilian casualties, particularly if urban warfare breaks out in the streets of Baghdad. But others contend that the use of chemical agents would violate the 1997 Chemical Weapons Convention (CWC).

These objections were delivered to President George W. Bush and British Prime Minister Tony Blair in a 20 March letter from interest groups including the Union of Concerned Scientists, and Physicians for Social Responsibility. The letter urged the two leaders to outlaw the use of riot-control agents or calmatives in Iraq.

The topic first surfaced on 5 February, when US defence secretary Donald Rumsfeld told the House Armed Services Committee that he was looking for ways to allow commanders in the field to use riot-control agents, and implied he would seek presidential approval for such measures.

The United States has a legal framework that could permit such action. After the Vietnam War, the then President, Gerald Ford, ruled that chemical riot-control agents, such as tear-gas, could be used in certain circumstances but only with permission from the White House. One such situation would be to disperse civilians that enemy troops were using as a human shield.

The White House has not said whether such permission has been sought for the current conflict. The Pentagon confirms that chemical smoke and pepper spray have been used in previous conflicts, but declines to say what has been supplied to units in Iraq.

Some observers contend that US forces

are carrying more dangerous caltative agents, comparable to the gas used to end a siege in a Moscow theatre last October that resulted in the deaths of over 120 hostages and 41 hostage-takers. "We can document Pentagon research on these agents," says Edward Hammond of the Sunshine Project, a pressure group based in Austin, Texas, that opposes the development of chemical weapons. "I think they are chomping at the bit to use these things," he says.

But the Pentagon denies that such a research programme exists. The Department of Defense "is not pursuing any chemically or biologically based incapacitating agents", says a Pentagon spokesman.

Experts also disagree on whether the use of riot-control agents would violate the CWC. The document's wording covers chemicals that cause "temporary incapacitation", but its real targets are strongly toxic substances such as nerve gases, says Jim Lewis, a senior fellow at the Center for Strategic and International Studies in Washington DC. "It was poor drafting," he says. "As it stands, it includes the mace on your key chain."

Furthermore, proponents of caltative agents argue that non-lethal gases save lives. If enemy soldiers are hiding in a building with civilians, for example, it is safer to drive everyone out than to go charging in, they point out.

Objectors counter that the Moscow incident showed that caltatives can be lethal (see *Nature* **420**, 7; 2002). And some take exception to the use of milder agents such as tear-gas. Mark Wheelis, a microbiologist at the University of California, Davis, says that many countries would see the use of tear-gas to disarm Iraq of chemical weapons as hypocritical. "If the United States uses riot-control agents, most of the world would consider it a violation of the CWC," he says. ■



Big issue: projects such as the Three Gorges Dam on China's Yangtze River continue to make waves.

## Conflicts undermine global forum's push for safe water

David Cyranoski, Kyoto

The organizers of last week's 3rd World Water Forum in Japan are saying that it succeeded in translating a vision for solving the global problems of water supply into action. But, upstaged by the war in Iraq and facing trenchant criticism from environmental groups, the forum struggled to win the recognition that its architects sought.

Some 24,000 people from 182 countries attended the event, held on 16–23 March in Osaka, Kyoto and nearby Shiga. By its close, 422 'water actions' — concrete measures by which different nations demonstrated their commitment to solving water problems — had been submitted.

But critics said that most of these actions — as well as agreement on a six-page statement signed by ministerial delegates to the meeting from all countries — were arranged in advance. This left many participants asking whether the mammoth event had actually furthered the effort to secure safe water supplies.

Host nation Japan set the tone for the meeting by promising to spend ¥16 billion (US\$133 million) this year to bring clean drinking water to people in poor countries, and to train 1,000 water-supply specialists over the next five years. It was one of several initiatives announced in pursuit of the United Nations' goal of halving the number of people without access to safe drinking water by 2015 (see *Nature* **422**, 251–256; 2003).

But the outbreak of war in Iraq took its toll on the meeting, with leading figures including Japanese foreign minister Yoriko Kawaguchi pulling out of the event. Reporters from across the world complained that their stories on the meeting were not finding space in news pages dominated by war coverage.

Speakers at the forum sought to align its goals with the international debate over the war. For example, former Japanese prime

minister Ryutaro Hashimoto, who chaired the forum's steering committee, pledged to send experts to supply water to those displaced by the war.

Environmental groups, meanwhile, attacked the forum's ministerial statement for skirting controversial issues. Agnes van Ardenne, minister for development cooperation in the Netherlands, criticized the document for its silence on problems related to climate change and the private ownership of water, and questioned the need for future water forums.

Hydrologists and environmentalists at the meeting clashed over the value of large dam projects. Mahmoud Abu-Zeid, president of the World Water Council and the forum's co-organizer, called for a "large increase in the number of dams". But environmentalists attacked the ministerial statement for failing to address problems related to dam construction. "This is a very weak document," said Ger Bergkamp, coordinator of the World Conservation Union's water and nature initiative. Abu-Zeid pointed out that the forum couldn't be expected to solve such problems overnight. "We did not promise a treaty," he said.

Organizers said that the forum had exceeded expectations in its main function of building links between scientists, officials and activists engaged in water issues, noting that more than 100 new commitments to collaborate had emerged.

In a statement summing up the forum's achievements, World Water Council vice-president William Cosgrove said that, despite the forum's conflicts, such commitments reflect attention to its key agreements — "that community-level public participation is fundamental to achieving the goals of the forum", and "that the common basic requirement for water is an opportunity for cooperation and peace". ■

## Mystery virus slow to yield its identity as patient numbers rise

Helen Pearson, New York

Health officials were intensifying their efforts this week to control an international outbreak of mystery pneumonia.

Officials at the World Health Organization (WHO) are reporting a continuing rise in their tally of patients who have contracted the condition — dubbed severe acute respiratory syndrome (SARS). As of 24 March, 456 cases and 17 deaths had been reported in 13 countries.

But doctors have partially succeeded in restricting the spread of SARS by isolating patients and using measures such as face-masks to block exposure to coughed-up droplets bearing the pathogen.

A worldwide group of at least 11 collaborating laboratories is also making headway in identifying the microorganism behind the disease, and developing a simple diagnostic test.

On 24 March, researchers at the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, reported that the SARS pathogen might be a new member of the coronavirus family, which includes viruses that cause some common colds and respiratory infections. Particles found in the tissues of some SARS patients had a similar structure and genetic sequence to known coronaviruses.

This followed suggestions last week that the pathogen responsible might be a previously unknown member of the Paramyxoviridae family, which includes the agents that cause measles and mumps.

But those involved in the investigation warn that these candidate viruses may simply be coincidental infections. Wolfgang Preiser at the Institute for Medical Virology in Frankfurt, Germany, says that it would be "premature to claim victory" in identifying the virus.

Preiser's laboratory, among others, is now attempting to grow the pathogen and match its genetic sequence with known specimens. But to prove beyond doubt that a candidate virus is causing SARS, additional 'gold-standard' tests will be needed, says virologist Albert Osterhaus of Erasmus University in Rotterdam, the Netherlands.

One such test is the enzyme-linked immunosorbent assay (ELISA), which can detect antibodies produced by the patient's immune system to fight a particular virus.

The CDC announced that it has already had some success with similar techniques and that several recovering patients seem to have antibodies against the coronavirus that they have isolated. ■



# Moderate flies flag for science on Capitol Hill

Geoff Brumfiel, Washington

The hundreds of researchers who trek to Washington each year to talk to the government about science and science funding don't always find Capitol Hill friendly territory. Most in Congress are businesspeople or lawyers by trade — only 3 of the 535 of them are qualified scientists.

But there's one place on Capitol Hill where science can rely on a sympathetic hearing: the Science Committee of the House of Representatives. And since 2001, under the chairmanship of Sherwood Boehlert (Republican, New York), the committee has been enjoying a new lease of life.

Bills that used to be ignored — such as a measure to double funding for the National Science Foundation (NSF) over five years — have been quietly moving through Boehlert's committee and then becoming law. The structure of the new Department of Homeland Security actually reflects Boehlert's demand for a top-level research administrator. And when the space shuttle Columbia fell tragically to Earth last month, the science committee took the lead in cajoling NASA to mount a fully impartial investigation.

Boehlert himself is a non-scientist whose moderate views and long-time embrace of environmental conservation place him firmly on the liberal edge of today's Republican Party. But he won the grudging respect of the party leadership for his tenacious defence of environmental regulation, and was allowed to succeed the fiery conservative James Sensenbrenner (Republican, Wisconsin) to take charge of the committee in January 2001.

Two years later, science lobbyists and Capitol Hill veterans credit Boehlert with raising the committee's profile and reputation, hiring dedicated PhDs as committee staff and making some progress in educating fellow members on current scientific issues. "He's made the science committee more relevant by far than it has been in 20 years," says Frank Cushing, a former member of staff with the House appropriations committee who now lobbies Congress.

Sitting in his high-ceilinged office, a relaxed Boehlert, sporting a tie bearing the colours of his local college basketball team, says that the committee is finally finding its voice. "In the weekly meetings between the House leadership and the committee chairs, the science committee's voice is heard, and often heeded," he says.

He adds that better relations with Senate committees have bolstered his committee's standing. "In the past, the House would do something and the Senate wouldn't respond, or the Senate would do something and the House wouldn't respond. Now we enjoy a very productive partnership."

Boehlert continues to dissent from some



A sympathetic ear: Sherwood Boehlert has won credit for reviving the House science committee.

of the House leadership's views on science — voting last month, for example, against a blanket ban on cloning research that was overwhelmingly passed by the House.

In keeping with the climate in the rest of the city, he has lately been devoting a lot of attention to security issues, including the thorny question of self-censorship in scientific publishing. Boehlert is unhappy with a Bush administration proposal, laid out by presidential science adviser John Marburger last October, that some research be classed as "sensitive but unclassified".

"It's just gibberish to say, 'sensitive but unclassified'," he says. "What does that tell a researcher?" He also doubts that self-censorship by scientists will work, and says that he supports a simple system in which research is either classified or unclassified. "That's the way I'm leaning — but we're going to have more hearings on it," he says.

## The final frontier

Boehlert's 47-member committee oversees most non-biomedical civilian research programmes in the United States, including those at the NSF, the Department of Energy and the Environmental Protection Agency. But its largest charge is NASA. In the wake of the Columbia accident, the committee has been monitoring the investigation — in partnership with the Senate commerce committee, chaired by John McCain (Republican, Arizona) — while rallying in support of the embattled manned space-flight programme and its centrepiece, the International Space Station.

Boehlert says he is worried about the ability of the pared-back space station to do research. "Most of us were operating under the assumption that it took two-and-a-half

people to operate the station. Now NASA says it can operate with two staff and still get some research done. Is it enough research to justify the cost? Not by a long shot." Yet he remains optimistic that the station, and manned flight in general, will prove beneficial in the long run. "I've got to be very pragmatic as I look at manned space flight in the future, and I'm going to have to ask, 'is it worth it?' My gut reaction right now is yes." Boehlert is planning further hearings to look at the future of NASA's manned space-flight programme.

A committee hearing this week will investigate the complaints of foreign scientists who are being delayed from entering the United States by new immigration laws. "What we desperately need in the twenty-first century is an increased emphasis on international research collaboration," Boehlert says. "It's short-sighted and counter-productive to close our borders."

He adds that much of his committee's time in the coming year will be devoted to keeping a healthy budget for science as lawmakers move to cut spending to pay for homeland security and the war with Iraq.

But the five-year NSF budget plan passed by Congress last year is only a recommendation, and Boehlert must still get Congress's appropriations committees to implement it. He says he plans to do this "by convincing them to the best of my ability that this investment is going to pay handsome dividends".

In the meantime, Boehlert's persuasiveness and middle-of-the-road manner continue to accrue respect for his committee in Congress. A glowing profile of the committee that appeared last week in *The Hill*, Capitol Hill's local newspaper, echoed the message he wants to project: "Quiet committee rising," the headline said. ■

M. CAVANAUGH

## Long-haul balloon makes more of a splash than planned

**Washington** It should have ushered in a new generation of ballooning. But after waiting almost two months for favourable weather at its launch site in central Australia, NASA's new Ultra Long Duration Balloon — designed to fly for about three months at a time — returned to Earth just 12 hours after its launch on 16 March.

The gigantic pumpkin-shaped balloon never inflated evenly, and it eventually lost pressure after taking off, says programme manager Steve Smith of NASA's Wallops Flight Facility in Virginia. After flight directors decided to bring it down, the balloon landed in a shallow pond. "It's the only body of water in 500 miles," joked Smith.

The scientific payload, designed to measure ultraviolet radiation from the Earth's atmosphere, appeared to be undamaged. But the balloon programme, which aims to place equipment almost completely above the atmosphere during 100-day flights (see *Nature* **421**, 308–309; 2003), is on hold while engineers fix the problem. An agreement to fly over Russian territory on flights across the Northern Hemisphere has also been delayed.

## US Senate upholds ban on oil drilling in wildlife refuge

**Washington** US environmental groups celebrated victory last week in the fight over oil drilling in Alaska's Arctic National Wildlife Refuge.

An amendment to a Senate budget bill, passed by 52 votes to 48 on 19 March, has ended current plans for oil exploration in the refuge. President George W. Bush's administration had pushed for drilling to be authorized, arguing that opening up new domestic oil supplies would reduce the United States' dependence on foreign sources. But environmentalists had campaigned against drilling in the refuge, on the grounds that this would damage the region's fragile ecosystem (see *Nature* **422**, 103; 2003).



Cold comfort: no oil from Alaska's Arctic refuge.

## Stargazer sees red

**Sydney** A British astronomer faces deportation from Australia after emblazoning Sydney Opera House with an anti-war slogan.

Armed with red paint and rollers, Will Saunders of the Anglo-Australian Observatory in Sydney and Australian environmentalist David Burgess painted "No war" on the tallest crest of the iconic building in protest over Australia's involvement in the Iraq conflict. Both were arrested upon their descent and charged with malicious damage.

Saunders was granted bail, but while telling reporters that he loves his adopted country and would help to clear up the graffiti, he was re-arrested at the behest of the Australian Immigration Department. He now faces the



possibility of being sent back to Britain if his work visa is revoked.

A cosmologist whose research concerns the large-scale structure of the Universe, Saunders has been working on a three-year visa at the observatory, an international facility that operates two optical telescopes.

**"We're thrilled that a bipartisan group of senators stood up to protect this spectacular landscape," says Carl Pope, executive director of the Sierra Club, a Washington-based environmental group.**

**The Bush administration could still attempt to pursue drilling plans, although it may struggle to win support after last week's vote.**

## 'Low risk' of infection from smuggled meat

**London** About 100 kilograms of meat infected with foot-and-mouth disease are illegally brought into Britain every year, according to a new risk assessment.

The study, produced by the government's Veterinary Laboratories Agency and independent risk-assessment consultants SafetyCraft, says that most meat is brought in by individuals for personal use. The authors estimate that the chances of the imports reaching farms are small and would result in one farm animal, on average, becoming infected every 130 years.

Four million animals were culled to contain the last foot-and-mouth outbreak in 2001, and the government was due to publish an action plan along with the assessment on 25 March. The new measures include increasing the number of detector dogs working with customs, and continued plans for intelligence-gathering and further risk assessments.

♦ [www.defra.gov.uk](http://www.defra.gov.uk)

## Ex-professor's gift boosts Californian university's funds

**San Diego** A former engineering professor at the University of California, San Diego (UCSD), has given his old employers a donation of \$110 million.

Irwin Jacobs, now chief executive of Qualcomm, a communications technology company based in San Diego, announced the

donation to the university's engineering school on 15 March. Jacobs and his wife Joan have already given \$23.4 million to UCSD, which named its engineering school after them. The gift includes \$10 million for faculty recruitment over the next five years, with the remainder being used to increase the engineering school's funding for staff, teaching and research support.

The donation ushers in the public phase of a UCSD campaign to raise \$1 billion for research, teaching and endowed chairs. This latest gift pushes the total pledged so far to \$445 million, university officials say.

## Hunts for neutrinos and hominids pay off

**Washington** Palaeontologist Michel Brunet and astrophysicist John Bahcall each found themselves \$1 million better off last week after scooping Dan David Prizes.

Brunet, of the University of Poitiers in France, has spent decades exploring fossils unearthed in Asia and Africa. He was the subject of worldwide media attention last summer when he and his team published details of a six- to seven-million-year-old fossil skull believed to come from the earliest known hominid (see M. Brunet *et al.* *Nature* **418**, 145–151; 2002).

His co-winner Bahcall, from the Institute for Advanced Study at Princeton University in New Jersey, has studied subjects including stellar evolution and dark matter, but is best known for his work on solar neutrinos. He helped to develop a solar-neutrino detector in the 1960s and worked on models that accounted for the apparent shortfall in neutrinos detected by these devices.

The prizes were set up by Dan David, chairman of photo-booth operators Photo-Me International. The winners are judged to have made an outstanding "scientific, technological, cultural or social impact on our world".

♦ [www.dandavidprize.com](http://www.dandavidprize.com)



# A plea for diversity

Joan — formerly Jonathan — Roughgarden rejects established evolutionary ideas about gender roles and sexuality. Everyone wants to discuss the parallels with her personal experience. But the science speaks for itself, she tells Virginia Gewin.



Joan Roughgarden has devoted herself to challenging the 'male's-eye view' of animal mating patterns.

Darwin's legacy is under attack. His fundamental theory of evolution by natural selection retains its status as the bedrock of modern biology, but a new generation of researchers is challenging his later ideas about evolutionary aspects of sex and gender. And this movement has just gained an eloquent spokesperson — albeit a somewhat reluctant one — in Joan Roughgarden, a theoretical ecologist at Stanford University in California.

Roughgarden is now putting the finishing touches to *Evolution's Rainbow*, a book that catalogues the enormous variety of gender roles and sexual behaviours that are present in the animal kingdom. Her thesis is that this diversity undermines the generality of Darwin's prevailing theory of 'sexual selection', with its underlying vision of ardent males vying for the favours of coy, choosy females. "I want to be fair to Darwin's memory, but give a voice to the shared discontent," she says.

Roughgarden's unusual personal journey is bound to generate unprecedented publicity for this evolutionary debate. Five years ago, at the age of 52, Roughgarden — then known as Jonathan — surprised colleagues by undergoing a sex change. Although she admits that her experiences have helped to shape her scientific interests, she's worried that her story will be sensationalized by an inquisitive media. Indeed, at points during the interviews for this article, these concerns led to some uncomfortable moments, as *Nature's* questions

probed more deeply into Roughgarden's personal territory than she was happy with.

So let's start with the scientific background, and the orthodoxy that Roughgarden wants to overturn. Darwin devised his theory of sexual selection to explain traits such as peacocks' tails and stags' antlers, which seem to handicap their possessors in the everyday struggle for survival. He argued that such adornments evolve because they confer a sexual advantage, either by allowing males to compete directly with one another to gain access to mates, or by giving them a means to impress females. Darwin's successors embellished the theory, cementing the view that the basic pattern is one of promiscuous, competitive males and picky females. This dichotomy, theorists argued, stems from the fact that eggs are more costly to produce than sperm, meaning that females must choose a mate carefully to ensure the best outlook for her investment.

## Female intuition

This leaves little room for the idea that females might, under some circumstances, be just as competitive and promiscuous as males. It also offers no explanation for homosexuality, which is seen by many of Darwin's heirs as a theoretically inconvenient aberration. By providing an atlas of departures from the stereotypes of sexual selection, *Evolution's Rainbow* is Roughgarden's counter-manifesto.

Some researchers — many of them

women — began documenting animal examples of female promiscuity as long ago as the 1970s. Many were simply ignored, or else marginalized as 'feminist' thinkers. "They weren't heard and couldn't find a place in the system," laments Roughgarden, who acknowledges that her own experiences of discrimination have driven her to rail against such injustices. "My experience has given me an edge and motivation to speak up that someone who has not been so stigmatized may not feel so keenly," she says.

Roughgarden traces the epiphany that led to *Evolution's Rainbow* back to her participation in a gay-rights parade in San Francisco, shortly before her transition. "Looking at all these people, I knew biology — my field — said these people were impossible," she says. The conclusion, for Roughgarden, was inescapable: "These people aren't wrong, it's the theory that's wrong."

So began Roughgarden's move into the arena of sexual selection, the latest in a series of jumps that have seen her work on subjects as diverse as the foraging behaviour of Caribbean lizards, the biological oceanography of barnacles, and ecological economics. As Roughgarden sees it, Darwin's theory offers no convincing explanation of phenomena such as the anatomy of female spotted hyenas (*Crocuta crocuta*), which have genitals that are outwardly indistinguishable from those of males. *Evolution's Rainbow* is, in part, the story of her search for the answers

to such biological condundrums.

Roughgarden argues that, when it comes to gender and sexuality, the distorting lens of sexual-selection theory has tricked evolutionary biologists into abandoning their usual rigour in trying to explain diversity. It has been too easy, she suggests, to dismiss certain phenomena as anomalies or mistakes, rather than trying to understand why they evolved.

Roughgarden believes that we will begin to understand nature's diversity of sexual and gender expression if we pay more attention to the social interactions between the animals involved. To explain traits such as the female spotted hyena's 'penis', Roughgarden proposes a theory of 'social inclusion'. In many species, she suggests, an animal's reproductive success is dependent on its membership of a social group that controls access to key resources. In such instances, preferences among group members for an arbitrary trait can cause it to evolve rapidly in a runaway manner.

### Select groupings

This 'social selection' is similar, in essence, to Darwin's concept of sexual selection by female choice. "The difference," explains Roughgarden, "is that the benefit conferred on an animal carrying a social-inclusionary trait springs from being incorporated into powerful same-sex groups, and not from being preferred during mate choice by the opposite sex." Roughgarden predicts that a female spotted hyena with a small 'penis' would be excluded from female groups, and would thus be denied the opportunity to reproduce. "Exclusion is the social equivalent of lethality — the strongest form of natural selection," she concludes.

These ideas were developed against the background of the public emergence of

Roughgarden's female persona. For many people, the influence of this experience on the ideas expressed in *Evolution's Rainbow* is intriguing. But Roughgarden fears that an excessive fascination with her personal circumstances could undermine the book's scientific message. Indeed, she turned her back on Princeton University Press, which was originally lined up to publish the book, after her editor asked her to draw explicit parallels between her life and what she observed in nature. She has now teamed up with the University of California Press, which intends to publish *Evolution's Rainbow* early next year.

Roughgarden's entry into the sexual-selection debate is welcomed by many in the field — not least because of her unique personal perspective. "She brings special sensitivity to aspects of social diversity that I'm not sensitive to," says Patty Gowaty of the University of Georgia in Athens, whose own work on deviations from the predictions of sexual-selection theory has, on occasions, been attacked for its feminist overtones. Stephen Shuster, an invertebrate zoologist at Northern Arizona University in Flagstaff, agrees. "You've got to include different perspectives," he argues.

But some scientists privately wonder if — whether she likes to admit it or not — Roughgarden's own experiences of social exclusion have biased her view of the natural world. Even the broadly supportive Shuster, who is open to the idea of social selection operating alongside sexual selection, feels that Roughgarden goes too far in attacking Darwin's theory. "She throws out a very healthy baby with some slightly soiled bathwater," he says.

Roughgarden is affronted by suggestions that she has allowed her objectivity to be undermined. Indeed, she sees this critique as yet another manifestation of social discrimination. The straight white men who

dominate studies of evolutionary biology are rarely required to defend themselves against such accusations, says Roughgarden, who believes that they, in fact, have a stronger case to answer than she does. For example, she argues that biologists' prejudices have frequently caused them to record mountings as being male on female — without verifying the sexes of the animals involved.

### Moving on

Roughgarden wants to put her old life as Jonathan behind her, rather than having it dissected in the context of her science. Her academic peers describe Joan as a much happier individual, but Roughgarden says only that she feels "more human" as a woman.

When it comes to the discrimination that she experienced at Stanford after her transition, however, Roughgarden is more forthcoming. One male colleague even asked her to avoid the temptation to "dumb down the science" as a woman, she says. Roughgarden also claims that she was asked to step down as head of Stanford's Earth Systems Program, the much-lauded interdisciplinary undergraduate course that she established in 1992. But Franklin Orr, who was then Stanford's dean of Earth sciences, asserts that Roughgarden decided to quit in 1999 after discussing the time commitment involved.

In some ways, Roughgarden sees herself as lucky. Many transgendered people are hounded out of their jobs. She attributes the fact that she is still at Stanford to the support of Condoleezza Rice, formerly the university's provost and now President George W. Bush's national-security adviser. But this good fortune was a double-edged sword: although many transgendered people can begin a completely new life, Roughgarden could not do so without abandoning her scientific career — Joan's academic reputation depends largely on Jonathan's distinguished publication record.

Perhaps unsurprisingly, given her experiences, Roughgarden is something of a political animal. She stood, unsuccessfully, for the office of supervisor of her district in San Francisco in 2000, and is active on diversity issues at Stanford. Even her science has a political dimension. Roughgarden is convinced that the vernacular of sexual-selection theory — which, for example, labels males that don't engage in head-on competition for mates as 'sneaky', or 'female mimics' — perniciously permeates our understanding of human sexuality. "It denies minorities their dignity," she claims.

Roughgarden hopes to replace such thinking with "a unifying theory" that will celebrate the diversity of sexual and gender expression, rather than repress it. When *Evolution's Rainbow* hits the shelves, we shall see whether the world is ready to embrace her vision.

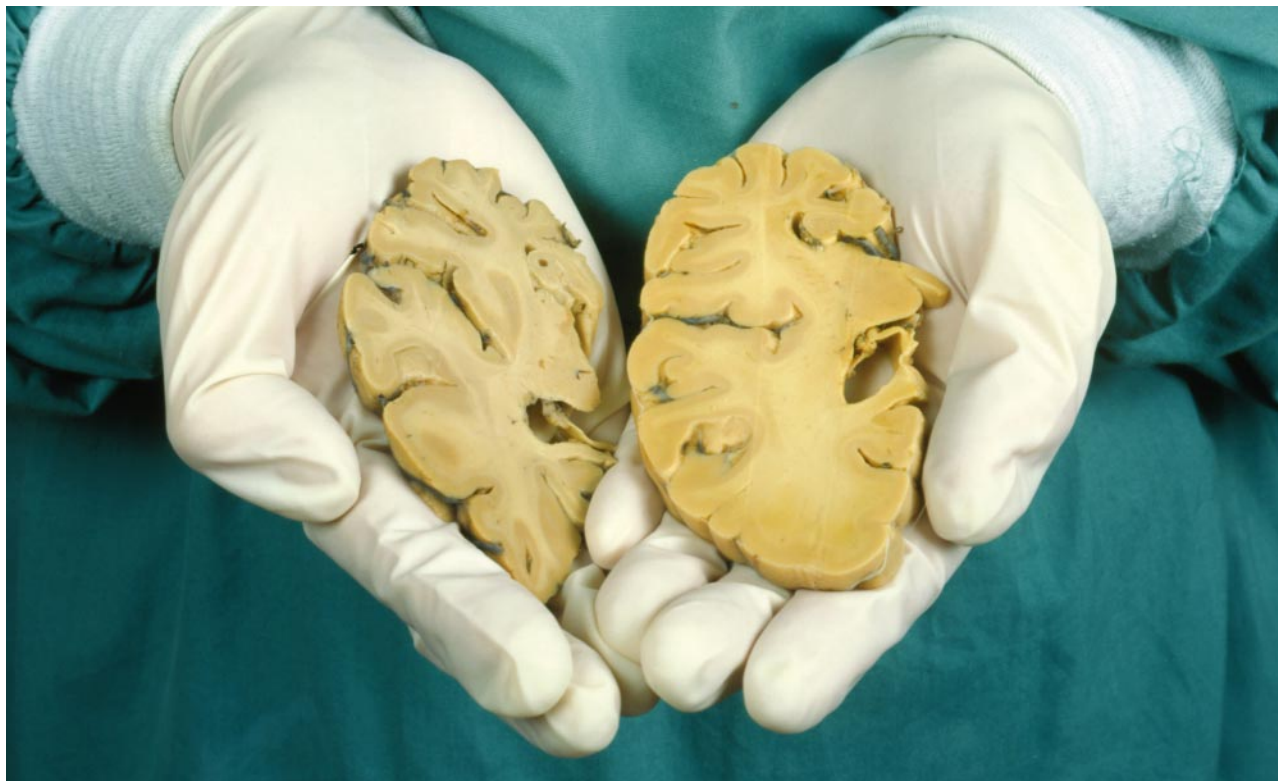
Virginia Gewin is a freelance writer in Corvallis, Oregon.



Making a splash: could a new theory of 'social inclusion' explain female spotted hyenas' unusual anatomy?

# Battle of the mind

Plans for a vaccine for Alzheimer's disease were derailed last year when clinical trials revealed serious side-effects. But as Erika Check finds out, this approach could be about to get back on track.



S. FRASER/MRC UNIT, NEWCASTLE GEN. HOSP/SPL

**Anatomy of dementia:** tissue damage and death in a brain from an Alzheimer's patient (left) leads memory loss and changes in personality.

Three years ago, a 72-year-old British woman with Alzheimer's disease became one of the first people in the world to test a promising experimental treatment for her condition. Starting in July 2000, doctors gave her and around 70 other patients doses of a vaccine designed to prompt their immune systems to attack and clear away the clumps of proteins that many believe cause the symptoms of Alzheimer's disease. Researchers were cautious, but hopes were high that the vaccine, which contained a form of the  $\beta$ -amyloid protein found in the plaques, could treat this distressing disease.

After promising results from this small safety trial, the pharmaceutical company behind the vaccine—Elan, in San Francisco, California—launched a larger trial the following year. But by the summer of 2001, problems began to emerge. The British woman was suffering from dizzy spells and

having trouble walking. In January 2002, Elan pulled the plug on its second trial after patients showed signs of brain inflammation. The woman from the original trial died of a blood clot a month later. Hopes for the vaccine seemed to be dead in the water.

## Take two

But not quite. The post mortem on the British woman, carried out by neuropathologist James Nicoll of the University of Southampton, UK, and published this month<sup>1</sup>, reveals that the treatment probably cleared some protein clumps. The first vaccine may have been unsafe, but many Alzheimer's researchers believe this approach could still produce a cure. Later this year, tests on a second-generation Alzheimer's vaccine could get under way. "Alzheimer's disease is devastating," says Dale Schenk, who led the Elan team that developed the

vaccine. "If you look at the data there's more evidence for this approach being potentially efficacious than most others."

The idea of developing an immunotherapeutic treatment for Alzheimer's emerged in the late 1990s. The notion was based on the amyloid hypothesis, the theory that misfolded clumps of a certain protein, known as  $\beta$ -amyloid, cause the symptoms of Alzheimer's. Elderly people commonly have clumps of  $\beta$ -amyloid, known as plaques, in their brains. But Alzheimer's sufferers have many more, and eliminating or modifying the plaques is now the focus of much Alzheimer's research.

In 1999, Schenk and his colleagues produced some of the earliest concrete evidence in favour of using the immune system to do this<sup>2</sup>. Working with transgenic mice that had been genetically modified to make human  $\beta$ -amyloid, they showed that these animals



developed fewer plaques if they were injected early in life with  $\beta$ -amyloid proteins. Even more intriguingly, they also found that injecting the protein cleared away plaques that had already formed in the brains of older mice.

Schenk and his colleagues thought that they knew roughly what was happening. They reasoned that the mice were treating the injected  $\beta$ -amyloid protein like an invading pathogen and generating antibodies against it. These antibodies then attached themselves to the  $\beta$ -amyloid and recruited other cells to eliminate the plaques. The cells they recruited — small immune-system cells known as microglia — chewed up and digested the plaques, leaving healthy brain cells behind.

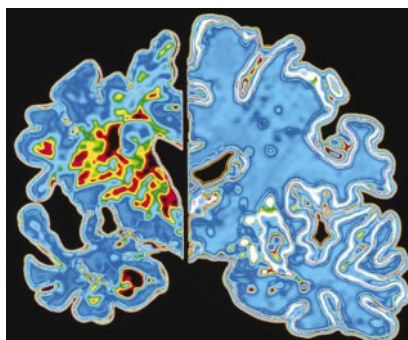
The technique, known as active vaccination, was considered enormously promising. “Everybody got excited, thinking this was going to be a cure,” says David Holtzman, a neuroscientist at Washington University in St Louis, Missouri. “They were rightfully excited — it was a very dramatic effect.”

### Antibody alternatives

Over the next three years, neuroscientists learned a lot about Alzheimer's vaccines. Two groups immunized similar transgenic mice with the  $\beta$ -amyloid and ran the animals through a water maze designed to test their memory skills. Vaccinated mice performed better than unvaccinated mice as the two groups aged<sup>3,4</sup>. The vaccine not only cleared plaques — it also seemed to be lessening the symptoms of the disease.

Meanwhile, a related immunotherapeutic approach was producing promising results at Elan's San Francisco labs. In 2000, Frédérique Bard and colleagues showed that rather than injecting mice with the  $\beta$ -amyloid protein, they could clear away brain plaques by injecting the animals with antibodies against the  $\beta$ -amyloid<sup>5</sup>. And last year, Holtzman and colleagues, together with a team from drug giant Eli Lilly in Indianapolis, Indiana, showed that this approach, known as passive vaccination, staves off learning and memory loss in transgenic mice that normally develop an Alzheimer's-like disease<sup>6</sup>.

The outlook for both approaches was rosy. None of the mouse studies reported side effects and Schenk says that unpublished primate studies had indicated that the  $\beta$ -amyloid-based vaccine was safe. Opting for the active approach — the better developed of the two — Elan started its first human safety trials with this vaccine in 2000, which included the woman later studied by Nicoll. Initial unpublished results from this trial were good enough for the US Food and Drug Administration (FDA) to give the go-ahead in 2001 for a larger trial of 375 sufferers in the United States and Europe, designed to find out whether the vaccine



Plaques multiply in the Alzheimer's brain (left).

would alleviate their symptoms.

But not everyone was convinced the trials were a good idea.  $\beta$ -amyloid collects in blood vessels in the brain, and an attack on these could cause brain haemorrhages. Another problem could come from a second line of immunological defences in which the immune system damages large areas around target proteins. That could be disastrous in the brain, as brain cells are unlikely to be replaced.

To make matters worse,  $\beta$ -amyloid is made from a larger precursor protein, which is sprinkled liberally throughout the brain. There was a theoretical chance that immune cells targeted at  $\beta$ -amyloid would recognize and destroy the precursor protein and any normal cells it was attached to. “You are vaccinating with something you already know is toxic, and which could create an autoimmune disease and inflammation,” says Blas Frangione, a neuroscientist at New York University. “We thought it was not a good idea.”

These doubts soon appeared justified. The second trial was halted when worrying signs of brain inflammation showed up in 6% of patients. And when the brain of the woman from the first trial was examined by Nicoll, T cells — an immune-system component that can cause inflammation — were found in the tissue surrounding her brain, suggesting that the vaccine had prompted a harmful immune response. Nicoll believes the same mechanism caused the inflammation in the larger trial.

But amid the bad news, evidence emerged that the vaccine was working. Last October, Roger Nitsch and colleagues at the University of Zurich in Switzerland showed how blood from 24 Swiss patients in the larger trial contained antibodies against  $\beta$ -amyloid<sup>7</sup>. Nicoll's autopsy also hinted that these antibodies might be clearing protein clumps, as that patient's brain lacked plaques in the temporal lobe, a region that is usually heavily affected in Alzheimer's patients.

Furthermore, the brain areas that lacked plaques also lacked certain kinds of abnormalities seen in brain cells near plaques, even though these areas were still marked by other characteristics of Alzheimer's disease, such

as  $\beta$ -amyloid in the blood vessels. Nicoll also found microglia, some containing fragments of  $\beta$ -amyloid, in the areas that lacked plaques.

For Schenk, the results are enough to suggest that the plaques were removed by the microglia. But other scientists are not so sure. The autopsy results have one major flaw: there was no way to examine the patient's brain before she received the vaccine, so it is impossible to prove that plaques were cleared.

There are other problems. One region of her brain — the frontal lobe — had the load of plaques usual in Alzheimer's patients. And Frangione says that a subset of Alzheimer's patients do not have plaques throughout their brains, and that the plaque distribution in the woman matched that of this subset. “We know there is a subgroup of patients who have very few plaques,” he says. “So the authors should not conclude that clearance of the plaques is due to the treatment.”

### Less is more

Nicoll disputes this, pointing out that the pattern of plaques is also similar to that seen in the mouse models. The issue is unlikely to be settled without more data. Elan may release results of the cognitive tests they ran on patients in the second trial later this year, and Schenk hopes to find some evidence of efficacy. But regardless of these results, there are enough hints that the vaccine may work for neuroscientists to want to continue with this approach.

The issue now is how to proceed safely. In the past year, the field has focused on two main options — modifying the  $\beta$ -amyloid protein used to make the vaccine, and running further studies of passive immunization.

The entire  $\beta$ -amyloid protein contains 42 amino acids, and Elan gave patients the whole protein. But the complete protein may not be needed to provoke an immune response. And because one end of the protein is thought to stimulate the antibody response and the other end the T-cell response, it may be that a vaccine that contained a small part of the protein would elicit a wholly helpful immune reaction.



Still optimistic: Dale Schenk believes that an effective Alzheimer's vaccine can be developed.

A. PASTERKAS/PL

Last year, neuroscientists Peter St George-Hyslop and Joanne McLaurin at the University of Toronto vaccinated transgenic mice with a small chunk of  $\beta$ -amyloid containing seven acids from the end of the protein that prompts the antibody response, and found that this generated antibodies that recognized the complete  $\beta$ -amyloid protein<sup>8</sup>. These antibodies destroyed  $\beta$ -amyloid clumps in a culture dish, indicating that they might recognize  $\beta$ -amyloid plaques in animals, although McLaurin has not yet assessed this.

But would such an approach be safer? McLaurin tested the kinds of immune response that mice launched against the short version of  $\beta$ -amyloid. She found that they responded to the vaccine by producing antibodies that are unlikely to activate T cells and that the animals did not produce the chemical signs that usually warn of inflammation. "All of this suggests that maybe by modifying the vaccine, you might be able to circumvent some of the inflammation," says McLaurin. Her group is now trying to find a small molecule that will mimic the effect of the limited chunk of  $\beta$ -amyloid. Small molecules are more attractive than macromolecules such as antibodies, as they are cheaper to make and easier to control.

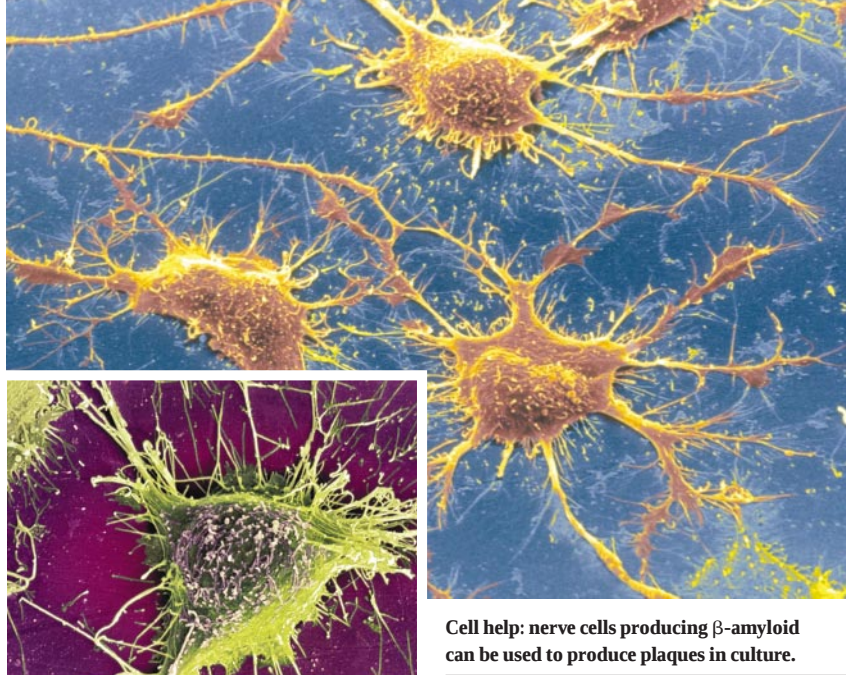
### Passive potential

Elan, in conjunction with the pharmaceutical company Wyeth of Madison, New Jersey, is also working on a vaccine based on a subunit of the  $\beta$ -amyloid protein. Even though the company was stung by the side-effects in its previous trial, Schenk says that it is important to go ahead with new active immunization trials, and Elan plans to ask the FDA for clearance to start new studies later this year.

Like last time, some researchers are advising against clinical trials. Giving a dose of vaccine that activates microglia in the brain is a dangerous proposition, argues Michael Mullan, director of the Roskamp Institute for Research into Neurodegenerative Diseases in Tampa, Florida. "It's like giving a baby a loaded gun," he says. "If you switch the microglia on, you're really not in control of what happens."

Concerns such as these are one of the reasons why scientists began experimenting with passive immunization. Now, many say that this option looks more attractive than the active approach. Elan wants to start a second trial later this year to test whether it is safe to inject antibodies directly into patients' bloodstreams. The reasoning is simple: the antibodies should cross into the patients' brains, then recognize  $\beta$ -amyloid, bind to it, and stimulate a microglial response — all without involving T cells.

But the story may be more complicated. It's not easy for large proteins such as antibodies to cross the blood-brain barrier. In



**Cell help:** nerve cells producing  $\beta$ -amyloid can be used to produce plaques in culture.

2001, Holtzman described how antibodies against  $\beta$ -amyloid may help clear away plaques without ever entering the brain. His team showed that injecting the antibodies into transgenic mice altered the levels of the protein circulating in the blood and the cerebrospinal fluid, which surrounds the brain<sup>9</sup>. This prompted Holtzman to develop his 'peripheral sink' hypothesis: if  $\beta$ -amyloid is constantly circulating from the brain out into the cerebrospinal fluid and blood, removing it from the blood would eventually cut levels in the brain. This could explain why passive vaccination was able to clear away brain plaques even though few antibodies are able to cross from the blood into the brain.

The hypothesis appears to be backed up by a study published this January by Karen Duff, a neurologist at the Nathan Kline Institute for Psychiatric Research in Orangeburg, New York. Duff and colleagues found that small molecules constructed to bind and remove  $\beta$ -amyloid from the bloodstream can clear the protein out of the brains of transgenic mice<sup>10</sup>. "This would mean we don't have to use antibodies any more," says Duff. "We could use a non-toxic binding compound to clear  $\beta$ -amyloid from the blood and we wouldn't even need to go into the brain, which is a big advantage to pharmaceutical companies."

Before such ideas can be evaluated, researchers will have to convince regulators that immunotherapeutic approaches are safe enough for clinical trials. The problems with Elan's earlier studies still cast a long shadow over Alzheimer's research. "The sad thing was that Elan used active immunization rather than passive, and many scientists had warned that this would be the result," Duff says. "I think the FDA will be looking much more carefully at any potential treatments in the future because this one had such disastrous results."

Passive immunization may appear a safer

bet, but there are hints that this approach may also have problems. Frangione points out that the human body could potentially respond to antibody injections by simply making antibodies against the antibodies. And researchers at University of Basel in Switzerland reported last year that passive vaccination of mice elicited small haemorrhages in blood vessels in the brain that were loaded with  $\beta$ -amyloid<sup>11</sup>. Nobody has yet tested passive immunization in people, so it's hard to say whether the same effect would occur.

If the regulators do give their approval, the trials should also provide the answer to the crucial question of whether the  $\beta$ -amyloid hypothesis, backed by many Alzheimer's researchers for the past decade, is in fact correct. Even if the vaccine proves to be unsafe, the results of the clinical trials could make or break the theory. If the plaques are cleared, but the distressing memory loss and cell death that mark the disease remain, neuroscientists will be left wondering what it all means.

"This is the ultimate test of the amyloid hypothesis," says Nicoll. "It's a very important question, and it looks like we might be able to get a handle on it in the next year." If Elan's trials show that its vaccine does not work, it could do more than end hopes for the immunotherapy approach — it might also send scientists back to the drawing board to come up with a new theory of Alzheimer's disease. ■

**Erika Check is Nature's Washington biomedical correspondent.**

1. Nicoll, J. A. R. *et al. Nature Med.* 10.1038/nm840 (2003).
2. Schenk, D. *et al. Nature* **400**, 173–177 (1999).
3. Janus, C. *et al. Nature* **408**, 979–982 (2000).
4. Morgan, D. *et al. Nature* **408**, 982–985 (2000).
5. Bard, F. *et al. Nature Med.* **6**, 916–919 (2000).
6. Dodart, J.-C. *Nature Neurosci.* **5**, 452–457 (2002).
7. Hock, C. *et al. Nature Med.* **8**, 1270–1275 (2002).
8. McLaurin, J. *et al. Nature Med.* **8**, 1–7, 1263–1269 (2002).
9. DeMattos, R. B. *et al. Proc. Natl Acad. Sci. USA* **98**, 8850–8855 (2001).
10. Matsuoka, Y. *et al. J. Neurosci.* **23**, 29–33 (2003).
11. Pfeifer, M. *et al. Science* **298**, 1379 (2002).



# Space: telescopes reveal the way forward

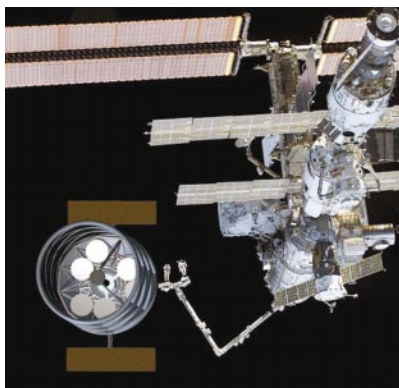
Astronauts' skills should be saved for major work in exploring the origins of the Universe.

*Sir* — Following the Columbia tragedy, scientists must re-evaluate the role of astronauts in scientific research (see *Nature* **421**, 559; 2003). We should ask astronauts to participate only in frontline research that would otherwise be prohibitively difficult. The continued exploration of the origins of the Universe and life itself with astronomical telescopes is such an area. For example, primitive life on an extrasolar planet could be detected as a chemical transformation of its atmosphere, as happened on Earth a billion or more years ago. But the telescopes must be of unprecedented size and sophistication. To be realistic, they will need continuing development in space, with controls and instruments upgraded as experience is developed.

Such evolution has been crucial to the success of the Hubble Space Telescope. Astronauts have carried out repairs and upgrades, transforming it into a long-lived and exceptionally productive observatory. Some advanced telescopes will build on this experience. The 10-metre X-ray Evolving Universe Spectroscopy (XEUS) telescope will be built at the International Space Station and operated nearby. A 10-metre successor to Hubble (see simulation in figure) for the optical and ultraviolet band could be similarly developed, eventually achieving the sensitivity to study extrasolar terrestrial planets.

Large infrared telescopes can also be extremely powerful, but present a special problem, because they need to be cryogenically cooled to be sensitive to faint heat sources. Thus they cannot be operated conveniently in low orbit where they would be warmed by Earth's radiated heat. The first large telescope of this type, the James Webb Space Telescope, is planned for operation a million miles away. But with a 6-metre aperture it will be much bigger than its cryogenic precursors, larger and more complex even than Hubble. If its parts wear out or need upgrading, or if its folded optics do not deploy correctly, it might be possible to bring the telescope back to a rendezvous as close as the Moon (*Nature* **419**, 666; 2002), but at considerable risk to astronauts. Otherwise we must launch an entirely new telescope at the same scale, or abandon the project.

For still larger and more complex cryogenic telescopes and interferometers, strategies to combine remote operation and more convenient astronaut access are clearly desirable. For example, a telescope could be assembled in low-Earth orbit as two linked spacecraft: a large but simple



Potential: a 10-metre successor to Hubble.

one housing the main mirrors and built to survive radiation damage; the other much smaller, lighter and containing the sensitive instruments and active optics controls. After verification of performance in orbit at room temperature, the two would be transferred independently to the remote operating point. The large structure would use efficient but slow solar-electric propulsion; the smaller instrument would be moved quickly through the radiation belts by conventional means. Both would carry enough fuel to allow a return to low-Earth orbit for repairs and upgrades by astronauts when necessary. Attachment and detachment could be achieved robotically, as with the unmanned supply

## Space: next step is an International Moon Base

*Sir* — Your far-sighted Editorial (*Nature* **421**, 559, 2003) on the future of human space exploration in the wake of the Columbia accident is a welcome contrast to some of your more sceptical comments about this subject in the past. Your call for vision, and clear endorsement of the exploratory value of human spaceflight, was a breath of fresh air. I would like to make two points.

First, you identify three possible future roles for humans in space: as geologists exploring near-Earth asteroids; as servicing crews for telescopes at L2; and as explorers on Mars. There are indeed good reasons for believing that all these activities, and more, would benefit from a human presence in space, but we must not forget the Moon.

The Apollo missions pioneered the use of astronauts as field geologists, and I shudder to think what the textbooks

vehicles that dock at the space station.

If astronauts again travel far from Earth, a manned lunar station would have great scientific potential and be a stepping-stone to a martian exploration. In general, the Moon is not an attractive telescope location, given its huge monthly thermal cycle. But the south lunar pole has attractive features for people and telescopes. Hydrogen is present, probably as water ice, and because the Moon's spin axis is not much tilted there are peaks experiencing uninterrupted sunshine.

A habitation on one of these could have constant solar power and a room-temperature environment. A telescope here would maintain constant deflection under gravity as it rotated once a month to track the stars. It could be built and tested in the sunshine, and for cryogenic operation would be cooled down simply by erecting a shallow bowl of multilayer insulation to screen out radiation from the Sun and Earth. For servicing or repairs, the bowl would be lowered to warm up the telescope. A 20-metre telescope built in this way would be of quite extraordinary sensitivity and flexibility, and would represent a huge accomplishment for lunar-based astronauts.

**Roger Angel**

Department of Astronomy and Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, Arizona 85721, USA

would now have to say about the early history of the Solar System had Apollo not taken place. Even today, one can scarcely attend a scientific meeting on the subject without seeing geochemical and isotope analyses of Apollo samples presented in one context or another. Yet Apollo, quite literally, only scratched the surface of the Moon, and there is so much more to learn. Moreover, while Mars does indeed beckon, we should probably learn to operate successfully on the Moon before attempting this much greater challenge.

Second, your Editorial implied that human space exploration is solely NASA's responsibility. But if such exploration is worthwhile, as I believe it is, we should all share in the costs and the risks as well as in the benefits. The pre-eminent value of the much-maligned International Space Station is not so much the science to be performed on it, valuable though that is, but the model it provides for international cooperation in space. We should aim to build on this experience to develop a



global human spaceflight infrastructure from which science can only benefit — an international Moon base being the obvious next step.

**Ian Crawford**

*School of Earth Sciences, Birkbeck College,  
Malet Street, London WC1E 7HX, UK*

### Illustration: database pictures tell a true story

Sir — Julio M. Ottino in his Commentary “Is a picture worth 1,000 words?” (*Nature* **421**, 474–476; 2003) divides images into two categories: those that convey data and those that illustrate scientific ideas. He defends the practice of image manipulation as sometimes being a necessary part of the process of discovery, yet expresses concern about the blurring of the line between fantasy and reality in scientific illustration.

It may be that the two categories are not that distinct. The European Space Agency's programme Innovative Technologies from Science Fiction for Space Applications assumes that even the most fantastic illustrations may be a useful stimulus to science. On the other hand, images that purport to convey factual data may convey something else entirely. Many observers, for example, were convinced that they could see little human beings in the sperm images produced by the sixteenth-century microscopist Antoni van Leeuwenhoek, and they recorded this observation as fact, presumably influenced by their beliefs.

Today, we can and should capture not just an image but information documenting the process of image creation itself, from the original unmodified data to the final web-ready or journal-ready artwork. If questions arise about the interpretation of an image, we need to be able to go back to the raw data or, if this is impossible, at least to have a full record of what was done to it, and why. For example, in performing video-enhanced contrast microscopy, one always subtracts a digital background ‘mottle’ image from the live video stream to obtain the mottle-free video frames that are recorded, and viewers need to be informed that background subtraction has been carried out. Similar arguments relate to the point spread functions used in preparing the deconvolved fluorescence images mentioned in the article.

Of course, as in the days of pen-and-ink illustration, scientists should still consider the purpose served by each image in their publications, and should make these objectives clear to their readers. Scientists today have the additional responsibility of

recording the processes by which images are created, so that these can be accurately replicated. Currently, such information is usually held only in the laboratory from which the image came, if it is recorded at all.

The BioImage Database Project (<http://www.bioimage.org>), part of ORIEL (Online Research Information Environment for the Life Sciences; <http://www.oriel.org>), will be a searchable database of multidimensional images of biological specimens. From its outset, we have felt it essential to acquire not just high-quality source images, but also the various images derived from them, and detailed metadata documenting the process of their creation. We believe that this approach addresses many of the issues raised by Ottino, and should be far more widely adopted.

**Chris Catton, David Shotton**

*Image Bioinformatics Laboratory,  
Department of Zoology, University of Oxford,  
South Parks Road, Oxford OX1 3PS, UK*

### Illustration: images fail to portray dynamic skies

Sir — An example of the visual deception practised by scientific illustrators seeking greater impact — as described by Julio M. Ottino in his Commentary (*Nature* **421**, 474; 2003) — is the increasingly prevalent practice of adding foreground stars and background galaxies to images generated from numerical simulations of galactic collisions. Examples appear in *National Geographic* **203**, 2, 2003; and on the Gadget website at [www.mpa-garching.mpg.de/gadget](http://www.mpa-garching.mpg.de/gadget). Another is one of my simulations, published unadulterated on the cover of *Nature* on 9 March 1989, which was later “improved” by the image-makers at NASA's Space Telescope Science Institute (<http://hubblesite.org/newscenter/archive/2001/22/video/c>) — without my consent.

Although one may agree that superimposed stars and galaxies add a bit of visual interest to the blankness of cyberspace, it is disturbing that these additions are entirely static. In reality, foreground stars hurtle past like snowflakes in a blizzard, and even background galaxies change noticeably over the hundreds of millions of years that are represented in the computer simulations. Static foreground stars and background galaxies undermine the basic idea of a dynamic Universe that these images and animations presumably attempt to convey.

**Joshua E. Barnes**

*Institute for Astronomy, University of Hawaii,  
Honolulu, Hawaii 96822, USA*

### Swift publication would reward good reviewers

Sir — I completely agree with T. Clausen and O. B. Nielsen, who say in Correspondence (*Nature* **421**, 689; 2003) that peer-reviewing needs to be adequately rewarded for the system to work efficiently. However, their proposed remedies, such as mandatory inclusion of reviews in CVs, do not seem promising, not least because this is already routinely done and does not seem to change the general picture.

While toiling on my considerable backlog of manuscripts to review, I came up with a scheme that might stand a better chance of improving the situation: diligent reviewing could be rewarded by speeding up publication of the reviewer's own research.

Currently, most of the prominent journals impose a strict deadline (typically between 7 and 14 days) on reviewers, yet the level of compliance is dismally low, as any author waiting for longer than two months for reviews of a submitted manuscript can testify.

My suggestion is that a researcher who meets the deadline for a particular journal with satisfactory reviews on, say, six consecutive occasions within a two-year period is guaranteed by the journal that his or her next submission will be given privileged status and reviewed within that same deadline. If the journal's editor cannot receive reviews of a privileged submission before the deadline, the journal would have to make a decision based on the other referees' comments or, if none had been received, accept the paper as it stands.

A possible variant is for journals that reject many submissions without review to guarantee to review a privileged submission (again, within the deadline). This is unlikely to burden these high-profile journals with piles of junk, because an individual writing six or more useful reviews for one of them within a limited time span is, by definition, a highly competent researcher.

Of course, a crucial aspect is the notion of a ‘useful’ review: to earn a privileged submission, reviews would have to be of high quality, as judged by the journal's editors, rather than one-liners. I believe that such direct feedback between a researcher's own publications and reviewing activity could seriously improve the peer-review system.

Feedback loops do wonders in biological systems; they just might work for science, too.

**Eugene V. Koonin**

*National Center for Biotechnology Information,  
National Library of Medicine, National Institutes of  
Health, Bethesda, Maryland 20894, USA*

# The price of collecting life

Overcoming the challenges involved in computerizing herbarium specimens.

Gideon F. Smith, Yolande Steenkamp, Ronell R. Klopper, Stefan J. Siebert & Trevor H. Arnold

The digitization of natural history collections in museums — the capture of data from specimen labels — is an essential first step in transforming these vast amounts of data into accessible, usable and useful information products. This process has attracted a lot of attention lately<sup>1</sup>, mainly because it is vital if these institutions are to remain relevant to their customers and stakeholders. However, very little has been done to accurately calculate the costs involved in such an exercise. What follows is a summary of our experience in southern Africa of computerizing herbarium (plant) specimens in the Southern African Botanical Diversity Network (SABONET) project. Our findings provide a valuable and accurate estimate of the costs involved for other, similar enterprises.

Computerization of natural history collections is an objective of the Global Biodiversity Information Facility (GBIF) ([www.gbif.org](http://www.gbif.org)), the All Species Project ([www.allspecies.org](http://www.allspecies.org)), and a goal of numerous major natural history collections and taxonomic institutions for more than three decades. The Botany Department of the Natural History Museum in London, for example, has designed and is currently testing an experimental database to quantify the time, effort and money spent on indexing natural history specimens in this way ([www.nhm.ac.uk/botany/cuttings/issue9/newsandviews/index.html](http://www.nhm.ac.uk/botany/cuttings/issue9/newsandviews/index.html)).

With access to comprehensive, interoperable databases, scientists and other users of taxonomic information will be able to rapidly compile biodiversity catalogues, study biogeographical patterns, and assist in both long-term conservation planning and crisis conservation management. In addition, we taxonomists should bear in mind that our salaries and project running costs are often paid from public funds. We should therefore be prepared to make our data accessible to interested and affected parties.

## Plants lead the way

Extensive work has already been done on animal collections, but the computerization of herbarium collections has mostly lagged behind. However, in southern Africa, the Pretoria National Herbarium Computerised Information System (PRECIS) database — one of the first of its kind and size in the world — has been under development since the early 1970s, providing the single largest

## Box 1 SABONET

SABONET ([www.sabonet.org](http://www.sabonet.org)) is a successful donor-funded, capacity-building project operating in Angola, Botswana, Lesotho, Malawi, Mozambique, Namibia, South Africa, Swaziland, Zambia and Zimbabwe. Botanists from these countries are involved in compiling specimen inventories, Red Data Lists and national plant checklists under the auspices of SABONET. Plant collecting expeditions target under-collected areas and under-collected taxa. The training of postgraduate systematics and biodiversity science students and herbarium personnel is also supported by SABONET. One of the project's

principal objectives is the "computerization of plant specimens stored in national and regional herbaria and botanical gardens".

### Projected outputs include:

- ✦ A checklist of the southern African flora, complementing ref. 3.
- ✦ A checklist of the endemic plants of southern Africa.
- ✦ A collated distribution map of the southern African flora indicating areas of high diversity and endemism.
- ✦ A comprehensive list of the grass taxa for all participating countries.

## Box 2 PRECIS

The PRECIS database consists of a number of components, two of which house the specimen records and the taxon names, respectively.

In the specimen sub-database, information such as family name, genus name, specific epithet, type status, identification level (degree of certainty of identification), temporary name, name of the person who identified the specimen, name of the collector(s) with collection number and the date of collection can be recorded. In addition, the region and country where collected, longitude and latitude, and a 1/4-degree grid reference<sup>4</sup> or exact location (GPS-reading) are also recorded. Habitat information (such as soil type, substrate, moisture regime, vegetation and biotic effects), and phenological information (such as presence

of flowers, fruit and seed, flower colour, growth form and height of plant) may also be recorded. There is also a 'notes' field to record any additional information on the plant's morphology, ecology, occurrence, common name, usefulness, toxicity, reproductive biology, horticulture, aroma and whether or not it is desirable (weed, invader, and so on). Finally, it can be noted if the specimen serves as a voucher for any particular study.

From the 'plant name' sub-database (taxon sub-database) one can obtain information on specific taxa, including synonyms, distribution within southern Africa and uses (for example, medicinal, food). Currently, only approved users have access to PRECIS, but the intention is for it to be available for anyone to use freely.

digital repository of botanical information in Africa. Since its small beginning, PRECIS has extended to ten southern African countries, largely under the auspices of SABONET (Box 1).

Although PRECIS can house a myriad of information (Box 2), the capture of information into selected fields has been emphasised for the past few years by SABONET. These selected fields include accepted name, synonyms and importantly, geo-referenced distribution data (GPS or grid location

data). This policy is in line with other major projects, including GBIF. GBIF focuses on species and specimen level data, and its DIGIT programme will concentrate on capturing and geo-referencing basic specimen label data that can later be extended to include other information.

It is easy to estimate the number of specimens housed in southern African herbaria. Following the publication of the comprehensive southern African Index herbarium and a few updates (ref. 2 and references therein), we know that 4,223,697 specimens are kept in the herbaria of the subcontinent (Fig. 1, overleaf). Since the early 1970s, 760,000 specimens kept in South Africa's national herbarium have been computerized — this figure being given a boost in the mid-1990s when SABONET extended the programme to nine other southern African countries. A further 374,532 specimens have been added to PRECIS since 1998, through the work of dedicated and

Several valuable lessons learnt by SABONET would be important for similar future projects.

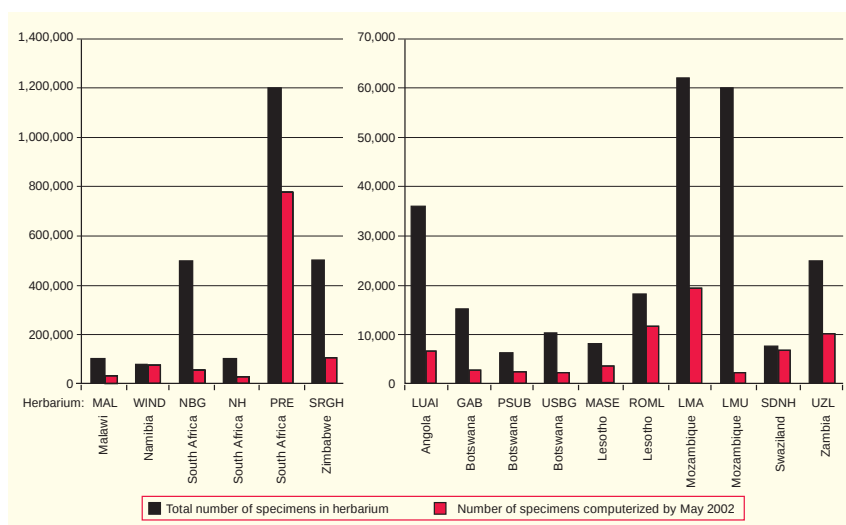


Figure 1 Total number of specimens housed in participating herbaria, as well as the number of specimens computerized by May 2002. All herbarium acronyms follow ref. 5.

part-time data capturers. A total of 1,134,532 of the specimens contained in southern African herbaria have therefore now been computerized (Fig. 1), and by extrapolation, between 481,250 and 604,681 specimens would ultimately be captured under SABONET.

From January 1998 to May 2002, it has cost an equivalent of US\$1,376,223 to computerize these herbarium collections in southern Africa (see Table 1 for a breakdown of the expenses). This figure includes aspects such as staff remuneration, consultants' fees, acquisition of computer hardware and software, field/collection trips, acquisition of equipment such as microscopes, herbarium cabinets, vehicles for fieldwork, workshops and training courses. During this period, 374,532 specimens were digitized — hence it costs about US\$3.67 per herbarium specimen in southern Africa. This figure may not exactly match costs elsewhere in the world — staff remuneration would be much higher in the developed world, but other expenses could be less, for example acquisition of microscopes, other laboratory equipment and computers, and the need for vehicles for fieldwork would not be as necessary or as extensive as in developing countries.

Very little comparative data are available

from other institutions or similar projects. The Smithsonian Institution estimated the costs involved in digitizing its collections, including staff remuneration and the upgrading of computer hardware and software, but not including training, collection and so on. This fell between US\$1.63 (cataloguing standards) and US\$12.50 (enhanced standards) per specimen, depending on the amount of data captured. Computerization under SABONET requires the capture of information (for example geo-referenced distribution data) beyond mere cataloguing standards, but not necessarily to the level of fully enhanced standards, thus making it difficult to compare these figures accurately.

### Lessons learnt

Several valuable lessons learnt by SABONET would be important in the consideration of similar future or ongoing projects.

The importance of capturing specimen data — for both the institution and the region — must be emphasised by the project

Table 1 Breakdown of the cost of computerizing herbarium specimens in southern Africa, in US dollars

| Description        | 1998 to May 2002 |
|--------------------|------------------|
| Salaries           | 532,936          |
| Evaluation         | 6,314            |
| Meetings/workshops | 70,020           |
| Travel expenses    | 50,527           |
| Consultation fees  | 41,233           |
| Training courses   | 84,552           |
| Field trips        | 35,792           |
| Running expenses   | 189,002          |
| Equipment          | 293,401          |
| Maintenance        | 7,701            |
| Publications       | 64,740           |
| Total              | 1,376,223        |

Figures are the portion of SABONET expenditure on computerization. Total SABONET expenditure in this period was US\$2,610,367. Specimen numbers from ref. 6.

management. When the participating institutions are not convinced of this importance, they do not see this activity as a primary objective, and concentrate resources elsewhere.

Building capacity takes time. Newly appointed data capturers, electronic networks and system managers do not become confident overnight, and follow-up training courses will invariably be required (and should be budgeted for).

Participating centres should be encouraged to make use of local computer service providers. Institutions have on occasion waited weeks for SABONET to fix their 'faulty' database, when in fact the problem was a technical one that could have been fixed within a day by a local service provider.

A quick-response call centre is a necessity. When a problem with hardware, software, database construction or data gathering is encountered in a relatively isolated participating centre, a mechanism must exist to address and rectify it rapidly — for example, through a call to a central help desk, or courier service to send equipment, such as hard drives, to centrally based technical staff.

Data capturers should be encouraged to remain in their positions for at least a year. Data capturer positions have a high turnover, and new clerks are often appointed long before the next database training course is scheduled. As knowledge transfer is often inadequate, clerks remain unproductive until they have attended a course. New, untrained data capturers also do not usually understand the terminology in the database, and subsequently waste time contemplating and adding new variables.

Finally, a quality-control mechanism must be in place. Untrained data capturers are often responsible for many wrongly encoded entries and misspelt information, which must be corrected later, wasting resources and time.

Gideon F. Smith, Yolande Steenkamp, Ronell R. Klopper and Trevor H. Arnold are at the National Botanical Institute of South Africa, Private Bag X101, Pretoria 0002, South Africa. Stefan J. Siebert is at the SABONET Regional Office, Private Bag X101, Pretoria 0002, South Africa. Author for correspondence: G.F.S. at [gfs@nibpre.nbi.ac.za](mailto:gfs@nibpre.nbi.ac.za).

- Godfray, H. C. J. *Nature* **417**, 17–19 (2002).
- Smith, G. F., Willis, C. K. & Behr, K. *SABONET News* **6**, 214–215 (2001).
- Arnold, T. H. & De Wet, B. C. (eds). *Mem. Bot. Surv. S. Afr.* **62**, 1–825 (1993).
- Edwards, D. & Leistner, O. *Mitt. Bot. Staatssamml. München* **10**, 501–509 (1971).
- Holmgren, P. K., Holmgren, N. H. & Barnett, L. C. *Index Herbariorum, Part 1: The Herbaria of the World*, 8th edn (New York Botanical Gardens, New York, 1990).
- Arnold, T. & Siebert, S. *SABONET News* **7**, 92–96 (2002).

### Acknowledgements.

SABONET thanks GEF-UNDP for funding the computerization process of the participating southern African herbaria. The authors also thank H. Snyman (National Botanical Institute) and A. Weitzman (Smithsonian Institution) for their help.



King Protea: the national flower of South Africa.



# Galileo points the way

Scientists should think big, but pay attention to the details.

## Galileo's Finger

by Peter Atkins

Oxford University Press: 2003. 380 pp. £20.

The US edition will be published in June (420 pp., \$30).

Frank Wilczek

In *Galileo's Finger*, Peter Atkins, a distinguished chemist and skilled writer, plays a game that I suspect all reflective scientists have indulged in: identifying the Ten Great Ideas. On the whole, he plays it very well. This is a charming and ambitious book that I would not hesitate to recommend as a gift for a young person on the threshold of a scientific career, or as the basis for a course or discussion group on general science. Of its kind, it is the best I have encountered. The specific criticisms that follow should be taken within that context.

The book has three underlying themes: exposition of science, promotion of scientific theses, and discussion of the scientific method. The bulk of the book falls under the first aspect, exposition. Atkins' 'ten best' list comprises biological evolution, DNA as the basis of heredity, atomism, conservation of energy, entropy, quantum theory, symmetry, Big Bang cosmology, relativity, and logic as the foundation of mathematics, leading on to universal machines and undecidability. Each of these topics gets a richly textured chapter of roughly 40 pages, touching on the history of the idea, its intellectual relatives and recent developments.

For example, the chapter on atoms begins with a discussion of pre-Socratic speculations on a universal substance and Aristotle's misconceptions, and the prescient formulations of Lucretius. There is then a nice account of the emergence of scientific chemistry with Antoine-Laurent Lavoisier, John Dalton's predictive atomism, and Stanislao Cannizzaro's convincing, accurate formulation. Next comes the theory of atomic structure based on quantum mechanics and wave functions, and the foundation it provides for the periodic table of elements. It includes a striking modern image of silicon atoms, made using a scanning tunnelling microscope, and well-thought-out illustrations of intricate atomic wave functions, as predicted using Erwin Schrödinger's equation. The discussion is self-contained and elementary, but honest and never patronizing. Justice is done to subtleties such as electron spin and the Pauli exclusion principle.

The final chapter, entitled "Arithmetic", touches on foundational issues and is rather different to the others. Its relationship to established natural science is much less clear.



Reaching for the stars: Galileo explains his vision in this painting by Felix Farra from 1873.

Rather, in this chapter Atkins seems to be building towards some intriguing but wild speculations (clearly labelled as such) that appear in the last few pages, around the theme: "There is therefore an intrinsic logical structure to the universe, which has the same structure as arithmetic... Our brains, and their product mathematics have exactly the same logical structure as the physical universe itself."

Well, maybe. But I'd have liked to see some acknowledgment of the obvious objection that the mathematical formalisms that have been most fruitful in science, involving real numbers, limits and locality much more than discrete arithmetic, seem on the face of it to be contrived and indirect from the point of view of abstract logic.

The wonderful image of Galileo's finger — specifically, the middle finger of his right hand, which is preserved in the Museo di Storia della Scienza in Florence, Italy — leads off the whole book, and introduces its methodological undercurrent. The vessel that contains the finger bears a remarkable inscription: "Do not look down upon the relic of a finger... By preparing a small piece of fragile glass it first dared a feat which long ago was beyond the powers of young Titans, who piled mountains high in a vain attempt to ascend to lofty citadels." This is the essence of Galileo's scientific method and, I trust, ours: fearless in conceptual daring, but humble in its respect for observation and facts.

Unfortunately, Atkins' own discussion of scientific method is marred by a weird blunder, which appears in various forms at different places in the book. The most egregious example occurs prominently in his concluding epilogue, in the book's penultimate paragraph: "Suppose that a future version of M-theory settles down into a form that predicts all known masses of the fundamental particles, all the values of the fundamental constants, and the structure of spacetime, but suggests absolutely no other experiment. It would not be falsifiable."

But of course such a theory would be falsifiable. All you'd need to do is measure the masses and fundamental constants more accurately than before, and check whether the more accurate determinations agree with theoretical predictions. (And by the way, M-theory in its present form is nowhere near predicting all these things; actually, it predicts none of them.)

Earlier, Atkins makes related, although more measured, comments about quarks: "I warned that I was preparing your mind for the possibility that science will have to modify its criteria of acceptability. That was in connection with quarks: quarks have not been seen, and perhaps cannot be seen, yet we are increasingly confident of their existence as so much that can be verified flows from it. That is verification by implication, rather than verification by experimentation: verification by hearsay,

rather than verification by direct experience.”

But quarks have been observed, using appropriate special-purpose microscopes, within protons (Jerome Friedman, Henry Kendall and Richard Taylor received the Nobel Prize in Physics for this work in 1990), and have even been visualized directly through the jets that they produce following high-energy collisions. They have been studied in great detail, and their behaviour is in harmony with a uniquely predictive and beautiful theory. True, quarks cannot be isolated and studied at leisure, but neither can, say, living dinosaurs, nor any number of other scientific constructs (including, until fairly recently, electrons). There is nothing in the study of quarks and gluons that requires any compromise of scientific method. Indeed, perhaps nowhere else in contemporary science is that method applied with such quantitative rigour.

In general, Atkins' discussion of recondite frontiers of fundamental physics is somewhat shaky on the distinction between what is known and what is speculated. He short-changes the former and revels in the latter. This would be only a minor annoyance, except that some of his major rhetorical points and conclusions, hinting that a revision of scientific method is or ought to be under way, hinge on this very discussion. ■

Frank Wilczek is at the Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA.

## The fat of the land

**The Hungry Gene: The Science of Fat and the Future of Thin**  
by Ellen Ruppel Shell

Atlantic Monthly Press/Atlantic Books:  
2002/2003. 294 pp. \$25/£17.99

Peter Kopelman

*“Nay sir, whatever maybe the quantity that a man eats, it is plain that if he is too fat, he has eaten more than he should.”*

Dr Samuel Johnson

“Obesity is the trillion-dollar disease,” according to the chief business officer of one of the world's fastest-growing biotechnology companies. He confesses to Ellen Ruppel Shell, the author of *The Hungry Gene*, that if it were his choice, he would focus his company's activities on obesity. Presumably Shell's incentive to follow this lead combined scientific intrigue and a potentially huge readership; she admits that “like most Americans of a certain age and sensibility, I am painfully conscious of my body weight”. What follows is a ‘tabloid’ (as opposed to ‘broadsheet’) analysis of the causes and consequences of increasing body weight and obesity on global society.



**Full fat:** mozzarella is the cheese used on pizzas, but the buffalo-milk version has a variety of uses.

The book begins as a detective story with the ‘victim’ — a patient with morbid obesity — undergoing gastric bypass surgery. The clues are explored through a historical perspective of corpulence across the centuries, which confirms that extreme obesity and medical concerns are dominant factors of the twentieth and twenty-first centuries. To tell the tale, Shell, a correspondent on *The Atlantic Monthly* and an associate professor and co-director of the Program in Science Journalism at Boston University in Massachusetts, cleverly interweaves science and historical fact.

She leads the reader to rodent models of obesity and the classical parabiosis experiments in which the circulations of lean and obese mice were linked; these suggested the existence of a circulating satiety factor that curbs eating. Jeffrey Friedman is introduced as a larger-than-life scientist (with “a tall man's gently sloping shoulders, he has a tangle of dark hair, a moustache, and hands big enough to palm a basketball”) whose laboratory identified the satiety factor that is released from adipose tissue and named it leptin.

Here the reader is initiated into scientific research conflict: according to Shell, colleagues were used for their advice and favour and then dropped when the final discovery was made, patented and sold to the highest bidder. Many scientists who spent much time with the author may rightly blush at her assertions — it is an unusual experience to read so openly about esteemed and active scientists.

The author follows the thesis that leptin deficiency underlies human obesity by describing the findings from Stephen O'Rahilly's group at Cambridge University. They identified a pair of young cousins with leptin deficiency, the elder of whom lost significant amounts of weight following

treatment with leptin injections. The book begins to lose pace after this chapter with the realization that the cloning of a ‘hungry gene’ is not proving to be the salvation of mankind from extreme corpulence.

What follows is a predictable onslaught on industry: the pharmaceutical industry for corrupting physicians and scientists who are involved in obesity research, and naturally the food industry. Nevertheless, it is to Shell's credit that she realizes that drugs are not the solution to obesity: “Weight loss drugs are an iffy proposition for everyone. Unlike leptin, which resolves a rare genetic defect, obesity drugs are not meant to treat a specific pathology. Rather they are designed to interfere with what is essentially a healthy, smoothly running system.” True to this, she describes the horrific consequences of the “rapid westernisation of unassuming cultures in the Pacific islands”. In Kosrae, it seems, most middle-aged adults are now condemned to develop type 2 diabetes, coronary heart disease and hypertension, and die young, all because the worst of the Western diet has displaced traditional eating.

In contrast, Shell describes David Barker's belief that much of the current problem results from intra-uterine nutrient deprivation. Barker builds a theory of increased prevalence of obesity, hypertension and diabetes in later life on studies of babies conceived during the Dutch famine in 1944–45. This hypothesis is supported by Barker's work on a British cohort born in Hertfordshire when food was more plentiful. Barker's theory remains plausibly compelling, but Shell addresses counter-arguments only in the bibliography. Intra-uterine nutrition undoubtedly represents just one facet of a series of complex issues that underlie today's pandemic of obesity.

The author concludes her pursuit of the cause of obesity with an onslaught on the

food industry, especially fast-food outlets. She is correct to identify them as one of many aetiological agents, but she fails to take account of modern family life in which parents are working and have too little time for family pastimes. She discusses physical activity, or rather the lack of it, but the book touches only briefly on the “couch potato society”.

*The Hungry Gene* is aimed at a lay readership and is an interesting read. The author has certainly worked hard to uncover scientific theories, and supports her arguments with an extensive bibliography that includes several scientific publications. She brings the book alive with her investigative-journalistic (docu-journal) style. However, much of the more controversial evidence comes from conversations with scientists and their colleagues, or from newspaper articles, rather than evidence-based scientific reports. Some may dispute the glamorization of the science of a modern-day societal problem with the implication, by its title, that its origin is entirely genetic. Nevertheless, the book does largely represent current scientific thinking.

The detective story that identifies the victim reveals many culprits: the author concludes that “science has taught us that the obesity pandemic is less a matter of individual differences than of societal pressures, and of the power of the institutions that impose them”. Perhaps a fairer title for the book, as suggested by Dr Johnson, would be “The Hungry Society” — but that might not sell. ■

Peter Kopelman is at Barts and the London, Queen Mary's School of Medicine and Dentistry, University of London, Turner Street, London E1 2AD, UK.

## Clinical science not to be sneezed at

### Cold Wars: The Fight Against the Common Cold

by David Tyrrell and Michael Fielder  
Oxford University Press: 2002. 268 pp.  
£17.99, \$24.95

Nigel Dimmock

Now rapidly fading from memory, the UK Medical Research Council's Common Cold Unit (CCU) was a uniquely British venture, where volunteer members of the public stayed for two weeks for virtually no payment as human guinea-pigs. Probably in no other country could such altruism be found. The volunteers (and staff) lived on site in some comfort in prefabricated, single-storey, converted hospital wards on a windy hillside outside Salisbury in rural southwest England. The volunteers, who often came again and again, were incorporated into double-blind, placebo-controlled clinical



Chilling out: volunteers at the Common Cold Unit were infected with cold virus in a bid to find a cure.

trials, and therein lay the uniqueness of the CCU: real people were placed in a totally objective experimental situation.

In the late 1940s and 1950s, volunteers were infected predominantly with viruses (extracts of nasal secretions from people with colds) that could not be grown in cultured cells. Later there were other investigations, including trials of antivirals (it was home to the first clinical evaluation of  $\alpha$ -interferon), the evaluation of psychological parameters such as susceptibility to infection, and the effect of colds on hand-to-eye skills. Today's scientists will look enviously at the timescale over which funding continued despite there being no substantial progress; the funding was largely due to the upbeat personality and influence with the Medical Research Council of the CCU's director at the time, Christopher Andrewes.

*Cold Wars* is the story of the CCU, as told by its director for its final 30 years, David Tyrrell, and co-author Michael Fielder. The authors give a thorough account of common colds from prehistory onwards. We are told of the establishment of Harvard Hospital in 1941 as a potential Second World War front-line facility by Harvard Medical School and the American Red Cross, and of its conversion after the war by Andrewes to the CCU.

These were the early days of cell culture, which had only recently been made practicable by the industrial-scale production of antibiotics. The problem of growing cold viruses was cracked in the late 1950s by a combination of logic (human viruses are most likely to grow in human cells, and at 33 °C, the temperature of the nasal cavities) and serendipity (a faulty batch of medium).



## The red parrot sketch

Edward Lear is best known for his nonsense verse, but he was also an accomplished artist. At the age of 19 he was keen to establish himself as a painter of natural history and published *Illustrations of the Family of Psittacidae, or Parrots*, a collection of 42 hand-coloured lithographic plates, including one of a red and yellow macaw (above). These were greeted with great enthusiasm and compared to the work of John James Audubon. This rare collection of ornithological plates has now been published in PDF format by Octavo ([www.octavo.com](http://www.octavo.com)).



It did not take long to show that cold viruses are physically related to poliovirus, but differ in the sensitivity of their infectivity to low pH. This made sense because cold viruses cause respiratory infections and do not have to withstand gastric acidity. The characterization of the viral genome — an infectious, single strand of RNA — was published in *Nature* (204, 792–794; 1966).

I was recruited to the CCU by Andrewes in 1961, to characterize the fundamental nature of cold (now rhino-) viruses. Being at the CCU was a steep learning curve for a young scientist — I was expected to master viral biochemistry and biophysics, and become familiar with clinical virology. In addition, the international reputation of the CCU ensured a stream of eminent visitors to rural England from countries around the world, from the United States to the Soviet Union. Each member of staff, even the newest recruit, was expected to tell these luminaries their latest findings. The world's press also paid an annual visit. Meanwhile Andrewes, who had retired and come to live nearby, often dropped in, timing his visits for tea and cucumber sandwiches, which were the norm in those days! Andrewes was a consummate raconteur who demanded an audience, but not reciprocity.

Does the book work? Yes, because Tyrrell and Fielder write sound but instantly understandable science. However, the core of the book is made up of Tyrrell's personal insights during the time that he presided over the CCU as the clinician-scientist. He was renowned for his hard work, boundless enthusiasm and encouragement to all those around him. Everyone who worked at the CCU could add his or her own reminiscences of what was a very interesting place at which to work and, for most of us, to live. Personally I owe Tyrrell an enormous debt for his sympathetic help with my PhD studies.

This is a history of science over the past 50 or so years — a science that is almost unrecognizable now — and an account of how a great many scientists, clinicians, lay staff and volunteers made a huge contribution to the understanding of the common cold, but sadly formulated no cure. But what of the demise of the CCU? We are told simply that it outlived its usefulness and was closed in 1990. It would have been good to know why. Was this just misguided short-termism, and if it had it become too expensive, why was no public-private partnership (with the pharmaceutical industry?) put in place to save such a valuable clinical-trial facility? The site is now a housing development; at its entrance on the Blandford road, a small plaque notes that here the CCU was once located. ■

Nigel Dimmock is in the Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK.

## Science in culture

### Time will tell

**DNA is used as a symbol of hopes and fears in the genetic revolution.**

Martin Kemp

A human figure, naked and male (as far as we can tell in the absence of genitals), turns his face and the palms of his hands to the heavens, like one of the ascending elect in Michelangelo's painting *Last Judgement*. Standing atop the kind of classical column reserved for military heroes, he adopts a space-filling pose reminiscent of Leonardo da Vinci's famous man tracing the figures of a square and circle with outstretched limbs. He is fringed by a halo of light that obeys none of the laws of terrestrial optics.

This was the image that appeared on the cover of *Time* magazine on 17 January 1994. It now features in the exhibition "Representations of the Double Helix", curated by Soraya de Chadarevian and Harmke Kamminga at the Whipple Museum of the History of Science in Cambridge, UK. The cover draws attention to an article inside on "The Genetic Revolution" by Philip Elmer-Dewitt, subtitled "New technology enables us to improve on nature. How far should we go?" Francis Collins of the US National Institutes of Health (now director of the National Human Genome Research Institute) was cited as comparing the challenge with splitting the atom or going to the Moon.

The centrality of DNA to the genetic quest is signalled by its spiralling presence in the figure's thorax and abdomen, reminiscent of a giant structural spine. It is assumed to be an instantly identifiable icon for *Time*'s international readership, recognized even when, as here, the helices twist in the wrong direction. It would be interesting to take a census of how often this reversal has occurred in the 50-year iconography of the great molecule.

We read on the cover that "New breakthroughs can cure diseases and save lives" — a promise that has been more hyped than realized — but the positive connotations are anxiously tempered by the question "how much should nature be engineered?" The eerie, night-time atmosphere of the cover evokes a science-fiction realm in which excitement and fear are mingled, *Frankenstein*-like.

Such is the fame of DNA that it has featured on at least three other *Time* covers over a 32-year span: 19 March 1971, 13 September 1999 and 17 February 2003 (in the US edition, but 3 March 2003 in Europe). The most recent of these shows a modern Adam and Eve standing wreathed in golden helices that branch aloft into a fruiting tree of life (or of sinful knowledge?). The lead article on this occasion, written by Nancy Gibbs, deals with "The Secret of Life", and aims to show how "Cracking the DNA code has changed how we live, heal, eat and imagine the future".

We take the familiarity of DNA so much for granted that it is difficult to see its high profile as exceptional. How many abbreviated chemical names mean anything in the broader public domain? H<sub>2</sub>O and O<sub>2</sub> are familiar enough, and CO<sub>2</sub> has gained prominence in relation to global warming. All three have oxygen in common, which seemed to hold the promise of the secret of life in the late eighteenth and early nineteenth centuries. Like AIDS, HIV and, perhaps, BSE, its acronymic identity has achieved independence from any widespread awareness of its full name or chemistry.

The triumphant story of DNA is marked by signposts that become selectively significant when we retrospectively know where history was taking us. As Robert Olby showed in his recent article in *Nature* (421, 402–405; 2003), the great majority of papers on DNA up to 1960 made no mention of James Watson and Francis Crick's structure (*Nature* 171, 737–738; 1953) — it remained something of a 'sleepers'. The laconic visual and verbal presentation of their paper, which was a factor in its slow rise to fame, seems astonishing in retrospect, but we would do well to remember Watson and Crick's acknowledgement that "The previously published X-ray data ... are insufficient for a rigorous test of our structure. So far as well can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results."

Martin Kemp is in the Department of the History of Art, University of Oxford, Oxford OX1 2BE, UK.



Cover star: the DNA double helix has become an icon of our times.

# Developing themes

*Summarize yourself in the form of a title of a paper in Nature.*

Mechanisms of merging vocation and avocation.

*What was your first experiment as a child (on pets/siblings/dead flies, and so on)?*

I was fascinated by the fontanelle on my baby sister's head. Each time that I tried to push it, I was rebuffed by my parents. Being a kid a little over four years old, I remember being both curious and terrified. Curious, because I wasn't allowed to go anywhere near my sister's head, and terrified because I couldn't imagine how something as hard as a head could start out so soft in places.

*Who has been the most important mentor in your career?*

Actually there are two: Farish Jenkins Jr and David Wake.

*What single scientific paper or talk changed your career path?*

Doug Melton gave a talk in the mid-1980s at Harvard when I was then a graduate student. His talk revealed the power of molecular approaches to developmental genetics and the way in which these techniques could be used to answer classic questions about the evolution of form. My heart raced, largely because I felt that I was hearing the future.

*What's your favourite conference destination, and why?*

Bozeman, Montana. It is close to major fossil localities, fly-fishing sites, and far enough away from other population centres to make for an intimate meeting.

*What would you have become, if not a scientist?*

I would have run a fly-fishing store in Bozeman, Montana.

*What was the worst/most memorable comment you ever received from a referee?*

I submitted a paper on early frog fossils to *Nature*. The fossils were not particularly well-preserved and my co-author and I struggled over a three-dimensional reconstruction of the beast. A reviewer lamented the fact that we "describe a roadkill only to have its ghost resurrected later in the paper".

*You have the audience in your hands, but some smart-alec asks you the killer question you have no idea how to answer. What's your favourite response?*

"That's clearly one of our major issues, one that we've struggled with for some time. Let's talk about this after the seminar." I wouldn't

back away, and I'd be real up front about the challenges that the question poses. You never know: someone else in the audience may have a way out of the problem.

*What book is currently on your bedside table?*

John Keegan's *The First World War*.

*Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?*

Friedrich Nietzsche, Fyodor Dostoyevsky and Søren Kierkegaard. I wouldn't even try to keep the conversation light.

*You are on a plane behind two students obviously going to the same conference, who start to talk about your work. What do you do?*

I would introduce myself to them before things got embarrassing for them or me.

*What one thing would you rescue from your burning laboratory?*

My first edition of William Bateson's *Materials for the Study of Variation*.

*What do you most dislike about having research published?*

The artificial sense of completeness that it engenders. Scientific discovery and questions are never-ending.

*The Internet is the bane of scientists' lives because...*

...it continually reminds us that the pace of discovery is so numbingly fast that none of us can keep up with it.

*Apart from the obvious great discoveries (natural selection, DNA and so on), what overlooked or underrated discovery really changed the science in which you work?*

The great conservation of regulatory genes. This discovery opened up a whole host of new approaches and new questions for genetics, development and evolution.

*Some cultures designate actual people as 'living treasures', for what they and their contributions have meant to their fellow humans. Whom would you choose for this honour and why?*

Ernst Mayr. His knowledge and energy have propelled the field for years. The breadth of his knowledge and his passion for evolutionary biology are exemplars for his more 'junior' colleagues (which now includes virtually all of us).

*Is there a 'tyranny of reductionism' in how scientists are trained today? That is, are students taught to look more into the workings of things and not enough at the 'big picture'?*



## Neil Shubin

Neil Shubin is professor and chair of the department of organismal biology and anatomy at the University of Chicago. His research merges palaeontology with developmental biology.

I don't see it this way. Students are forced to be reductionist by virtue of the techniques that they must master. The resurgence of comparative and phylogenetic biology, through the fields of genomics and evolutionary development, have really forced a certain level of big-picture thinking on our students.

*What's the one thing about science that you wish the public understood better?*

That the appearance of design in organisms is the result of evolution. That science is a process of discovery, not a stale body of facts or laws.

*If you could direct more government funding into one area of science, where would you put it?* Geomorphology, in particular to understanding how the evolution of climates and landforms are linked. The keys to many ecological, evolutionary and conservation issues may lie here.

*Name one extravagance you can now get away with because of your eminence.*

I feel that I can ask any seminar question that I like. Even the most stupid or ignorant questions carry a profundity to them. I once asked a question that revealed that I knew little about the background of the speaker's talk. Most of the attendees assumed that I was asking something truly profound. I hope.

*Harry Potter or Lord of the Rings?*  
Harry.

# It's all relative

Elizabeth A. Kellogg

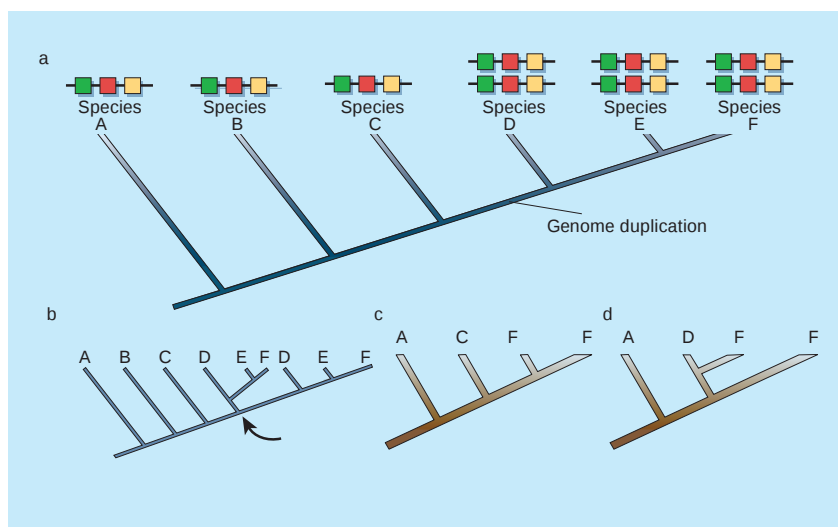
By constructing evolutionary trees of genes, researchers have detected three big genome duplications in the history of the plant *Arabidopsis*, and one in the recent history of yeast.

The final battle of the War of 1812 between Britain and the United States was the Battle of New Orleans. It was fought on 8 January 1815, but this date by itself is not particularly informative (and most students don't remember it). What is instructive is that the battle took place after the Treaty of Ghent, which ended the war — so the timing shows that the battle had no effect on the conclusion of the war. In other words, the relative time of the battle (before or after the treaty) is more interesting to the historian than is the absolute date. Two groups have now applied similar logic to identifying and dating gene-duplication events in yeast<sup>1</sup> and in the flowering plant *Arabidopsis thaliana* (page 433 of this issue<sup>2</sup>). Rather than trying to pin down the precise number of years since the last big duplication event, the authors instead used evolutionary trees to work out the relative time, before or after other events in evolutionary history.

Organisms frequently duplicate their genes. This can happen during the production of sex cells (meiosis), when segments of one chromosome are frequently swapped with the complementary segments in its matching pair; errors in this process can lead to the copying of one or a few genes at a time. The same outcome may result from transposition, where copies of genes are inserted in other parts of the genome. Alternatively, every gene in the genome can be copied simultaneously, in a single wholesale duplication event. This latter case is particularly common in flowering plants, but is thought to have happened in animals<sup>3,4</sup> and fungi<sup>5</sup> as well. Hybridization — mating between two species — frequently precedes whole-genome duplication in plants.

Duplicate genes provide raw material for gene diversification; an extra gene copy can acquire a new function, or the two copies can specialize by dividing up the ancestral functions. But this, together with the fact that extraneous genes are gradually lost over time, can make detection of the duplication event difficult. By analysing the evolutionary history of genes, however, Bowers *et al.*<sup>2</sup> have found at least three whole-genome duplications in the history of *Arabidopsis*, while Langkjaer *et al.*<sup>1</sup> document at least one major event in the recent history of the yeast *Saccharomyces cerevisiae*.

Most previous studies have assessed



**Figure 1** Effect of genome duplication on the history of individual genes. **a**, The evolutionary history of six hypothetical species, A–F. An ancestral species contained one of these genes, the green, red and yellow genes, and gave rise to species A–C, which contain one of each of these genes. Genome duplication later occurred in the ancestral species that gave rise to D, E and F. **b**, An evolutionary tree based on any one gene will show the duplication as a branch point (arrow) leading to two lines of descent, one to each copy of the gene. The tree will be the same for every gene in the genome. **c**, If a particular gene is sequenced only from species A, C and F, the two copies from F will appear more closely related to each other than to the gene from species C; so the duplication must have occurred after C and F diverged. **d**, An evolutionary tree that includes only one gene of a duplicate pair may still permit the relative time of duplication to be determined. In the tree shown, sequences were obtained for the one copy in species A, the two copies in F and one of the two copies in D. One copy from F appears more closely related to the gene from D than to the other copy from F; so the duplication in F must have occurred before D and F diverged.

genome duplication by using absolute time, analogous to trying to find the exact date of a battle. The average mutation rate of two genes that are assumed to have originated by duplication is estimated by counting the differences in base pairs — the building-blocks of genes — between the two, with each mutation interpreted as a tick of a molecular clock. The mutation rate speeds up, slows down and varies among genes, so it is clock-like only on average, but the intrinsic variability can be accommodated statistically; estimated rates and times of mutation for each gene can have standard errors attached to them. If the estimated divergence times of all gene pairs in a species fall into a single broad group within acceptable confidence limits, a genome duplication is inferred.

Many interesting evolutionary questions are raised by knowing the time of genome duplications relative to other events in evolutionary or geological time. So the molecular clock is generally calibrated to

determine how many years are represented by the average mutation. Calibration involves comparing genes in two species, and estimating the time since their common ancestor.

The problem is that the variance of the mutation process is modest compared with the uncertainty introduced by calibration. The date of the common ancestor of two species is estimated from a fossil, or from a geological event such as the separation of two continents. For example, the common ancestor of maize and rice must have lived somewhat after the origin of the grass family, for which the earliest fossils are pollen grains. Pollen that is unquestionably from a grass occurs in sediments dated to 55 million years ago, but there are also grains that could be grasses in rocks that are about 70 million years old<sup>6</sup>. If the maize/rice ancestor lived after that, then a reasonable date might be 50 or 60 million years ago. The calibration of the clock, in other words, includes what can only



be called a fudge factor, in this case of 10 or 15 million years — close to 20%.

The precise date of major genome duplications (measured by a molecular clock) can thus be compared with major events in evolutionary history (generally measured by a different molecular clock), using one or more calibration points (fossils). The error of the estimate is high, so correlations are difficult, if not impossible, to demonstrate rigorously.

Langkjaer *et al.*<sup>1</sup> and Bowers *et al.*<sup>2</sup> circumvent this problem by using relative time. Bowers *et al.* compare pairs of genes in *Arabidopsis* with those in cabbage (*Brassica*; from the same family), cotton (from a different family), pine (a seed plant, but not a flowering plant) and moss (a very distant relative), and for each gene they compute an evolutionary tree — the gene's pedigree. From the pattern of the evolutionary tree, they can determine when a duplication occurred relative to the evolutionary origin of other species (Fig. 1). The evolutionary tree (see Fig. 2b on page 436) shows a clear duplication event, affecting many genes in the genome, that occurred before the *Brassica/Arabidopsis* split, and before the members of the family Brassicaceae started to diverge. Similarly, Langkjaer *et al.* show that the yeast genome was duplicated before the divergence of *Saccharomyces* and *Kluyveromyces*.

After duplication, one copy of many of the genes in a duplicated genome segment is lost. Once duplicate segments have been identified, comparisons between the two allow the gene composition of the common ancestor to be estimated (Fig. 2). Having done this, duplicated regions that are even more ancient become apparent — pairs of genes and gene regions that were not initially identified because too many puzzle pieces were missing. At the same time, it is possible to identify the pattern and relative rate of gene loss. Repeating their evolutionary analysis for the newly identified duplicated segments, Bowers *et al.* were able to identify a more ancestral duplication event early in the evolution of the flowering plants, after the

ancestor of cotton and *Arabidopsis* (which are both dicotyledonous plants) diverged from the ancestor of rice and maize (which are monocotyledons). Another round of analyses revealed a duplication that was still more ancient, possibly occurring before the origin of the seed plants.

A historian, trying to dissect cause and effect, needs to know the relative times of battles and treaties. Similarly, the biologist needs to know the relative times of gene duplications, speciation events, major species diversifications, and events of Earth history. Approaches that involve the construction of evolutionary trees are designed specifically to assess relative time. Incorporating such an approach into future genome studies will undoubtedly lead to a clearer picture of the role of gene and genome duplication in the evolutionary process. By

increasingly dense sampling of evolutionary trees, even without complete genome sequences for every species, it is possible to distinguish single-gene duplications from whole-genome duplication. So the approach holds the promise of dissecting the dynamic processes by which genes and genomes evolve. ■

Elizabeth A. Kellogg is in the Department of Biology, University of Missouri-St Louis, 8001 Natural Bridge Road, St Louis, Missouri 63121, USA.  
e-mail: kellogge@msx.umsu.edu

1. Langkjaer, R. B., Clifton, P. F., Johnston, M. & Piskur, J. *Nature* **421**, 848–852 (2003).
2. Bowers, J. E., Chapman, B. A., Rong, J. & Paterson, A. H. *Nature* **422**, 433–438 (2003).
3. Gu, X., Wang, Y. & Gu, J. *Nature Genet.* **31**, 205–209 (2002).
4. McLysaght, A., Hokamp, K. & Wolfe, K. H. *Nature Genet.* **31**, 200–204 (2002).
5. Wolfe, K. H. & Shields, D. C. *Nature* **387**, 708–713 (1997).
6. Jacobs, B. F., Kingston, J. D. & Jacobs, L. L. *Ann. Missouri Bot. Garden* **86**, 590–643 (1999).

### Nonlinear dynamics

## Synchronization from chaos

Peter Ashwin

It isn't easy to create a semblance of order in interconnected dynamical systems. But a mathematical tool could be the means to synchronize systems more effectively — and keep chaos at bay.

Chaos and control are often seen as opposite poles of the spectrum. But the theory of how to control dynamical chaos is evolving, and, in *Physical Review Letters*, Wei, Zhan and Lai<sup>1</sup> present a welcome contribution.

Chaos is a feature in all sciences: from lasers and meteorological systems, to chemical reactions (such as the Belousov–Zhabotinski reaction) and the biology of living organisms. In most deterministic dynamical systems that display chaotic behaviour, selecting the initial conditions carefully can drive the system along a trajectory towards much simpler dynamics, such as equilibrium or periodic behaviour. But sensitive dependence on initial conditions — the well-known 'butterfly effect' — and the effects of noise in the system mean that in practice this is not so easy to do.

The aim of chaos control is to be able to perturb chaotic systems so as to 'remove' or at least 'control' the chaos. For example, in a spatially extended system, the aim may be to achieve regular temporal and/or spatial behaviour. Techniques introduced<sup>2,3</sup> and developed by several researchers over the past decade have sought to make unstable behaviour robust against both noise and uncertainties in initial conditions by stabilizing the system (using feedback<sup>3</sup>, for instance) close to dynamically unstable trajectories. These techniques have been very successful in controlling chaos, at least for low-dimensional systems.

Synchronization is a good example of a

chaos-control problem: synchronizing an array of coupled (interdependent) systems — such as the coherent power-output from an array of lasers — is of interest for technological applications. In biology, synchronization of coupled systems is a commonly used model<sup>4</sup>, and the presence, absence or degree of synchronization can be an important part of the function or dysfunction of a biological system. For example, epileptic seizures are associated with a state of the brain in which too many neurons are synchronized for the brain to function correctly.

In the simplest case, synchronization of two identical coupled systems (such as periodic oscillators) can be achieved through their coupling as long as it is strong enough to overcome the divergence of trajectories within either individual system. The required strength is indicated by the most positive Lyapunov exponent of the system: a Lyapunov exponent is an exponential rate of convergence or divergence of trajectories of a dynamical system, and the most positive Lyapunov exponent measures the fastest possible rate of divergence of trajectories. In particular, the fact that the individual systems have chaotic dynamics before they are coupled together means that the most positive Lyapunov exponent is greater than zero, and there is always a threshold below which synchronization cannot be achieved.

Synchronization in more general arrays can be done similarly, although with local coupling this can only be achieved with a

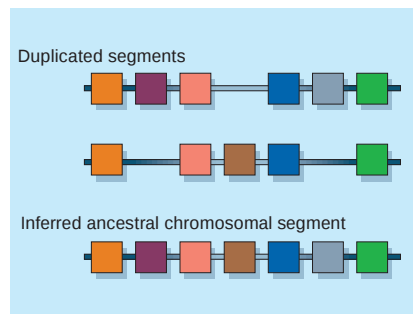


Figure 2 Duplicated chromosomal segments, showing some gene pairs. This pattern of duplication suggests that all seven genes may have been present and in the same order in the common ancestor.

coupling strength that grows with system size. This synchronization can be achieved without forcing the dynamics to become, for example, periodic. Hence, the problem of spatial control of coupled dynamics, although it still involves stabilizing dynamics that are inherently unstable, is easier to achieve than control of chaotic into simple dynamics. Control of synchronization can usually be achieved by careful design of the coupling, rather than resorting to feedback techniques. What then remains is to try to minimize the level of coupling required to achieve synchronization.

This is the problem that Wei, Zhan and Lai<sup>1</sup> have tackled. They have come up with a novel way of reducing the necessary coupling in an array by using wavelet decomposition of the matrix of coupling coefficients. Wavelets are mathematical functions that have been developed over the past decade or so as a powerful tool for signal-processing and numerical analysis. Wavelet analysis involves reducing a signal into a series of coefficients that can be manipulated, analysed or used to reconstruct the signal. Wei *et al.* make a small change to the low-frequency components in the wavelet-transformed matrix, before applying an inverse transform to obtain a modified coupling matrix. This turns out to be an efficient strategy for achieving synchronization at much lower coupling strengths.

Wei *et al.* test their method by synchronizing a ring of coupled Lorenz systems. The Lorenz system is a set of three nonlinear differential equations showing chaotic behaviour. In this proof-of-principle, a ring of Lorenz systems are coupled together linearly, their relations to each other represented by a matrix of coupling coefficients. A small change in this matrix (less than 2% for 64 coupled systems), through the wavelet transform, produces a much lower threshold of coupling to achieve synchronization. The authors show that their technique is robust even if the symmetry of nearest-neighbour coupling is broken.

It will be interesting to see if this method can be extended to more general arrays of coupled systems, to better understand control of spatial patterns. It may be that the work by Wei *et al.*<sup>1</sup> will suggest new techniques and structures for the design of local and global coupling in such systems. ■

Peter Ashwin is in the School of Mathematical Sciences, University of Exeter, Exeter EX4 4QE, UK. e-mail: P.Ashwin@ex.ac.uk

1. Wei, G. W., Zhan, M. & Lai, C.-H. *Phys. Rev. Lett.* **89**, 284103 (2002).
2. Ott, E., Grebogi, C. & Yorke, J. A. *Phys. Rev. Lett.* **64**, 1196–1199 (1990).
3. Pyragas, K. *Phys. Lett. A* **170**, 421–428 (1992).
4. Pikovsky, A., Rosenblum, M. & Kurths, J. *Synchronization: A Universal Concept in Nonlinear Sciences* (Cambridge Univ. Press, 2001).

## Neurobiology

# Ballads of a protein quartet

Mark P. Mattson

The fate of neurons in the developing brain and in Alzheimer's disease may lie with a four-protein complex that regulates the cleavage of two molecules spanning the cell membrane. The role of each protein is now being unveiled.

Scientific discoveries often originate in surprising places. Some years ago, for instance, researchers looking at how the brain develops received help from an unexpected quarter: studies of patients with Alzheimer's disease. This disease is characterized in part by the abnormal accumulation, in the brain, of a protein called amyloid  $\beta$ -peptide ( $A\beta$ ), which is a fragment of a larger protein, the amyloid precursor protein (APP), that sits across the outer membrane of nerve cells. Two enzymatic activities are involved in precisely snipping APP to produce  $A\beta$ , which is then shed into the brain. Curiously, one of these activities — dubbed  $\gamma$ -secretase<sup>1</sup> — was later discovered also to cleave Notch, a receptor protein that lies on the cell surface, and thereby to affect the way in which Notch regulates gene expression during normal development<sup>2</sup>. On page 438 of this issue, Takasugi and colleagues<sup>3</sup> add to our understanding of how APP and Notch are processed. Using genes and cells from flies

and humans, and the powerful new technology of RNA interference, these authors establish specific roles for four different proteins underlying  $\gamma$ -secretase activity.

For many years, much of the research into Alzheimer's disease has concentrated on identifying and characterizing the protein (or proteins) that generate  $A\beta$ . In the first step of this process, APP is cleaved at a specific point by a so-called  $\beta$ -secretase activity; the protein responsible for this activity was identified some four years ago. Cleavage by the  $\gamma$ -secretase activity then produces  $A\beta$  — but here the molecules at fault have been harder to pin down. An early hint came from the finding that mutations in a gene encoding the presenilin-1 protein occur in several families with inherited Alzheimer's disease; it was quickly shown that these mutations cause increased cleavage of APP to produce  $A\beta$ . So presenilin-1 was assumed to be the  $\gamma$ -secretase.

A surprising link to brain development

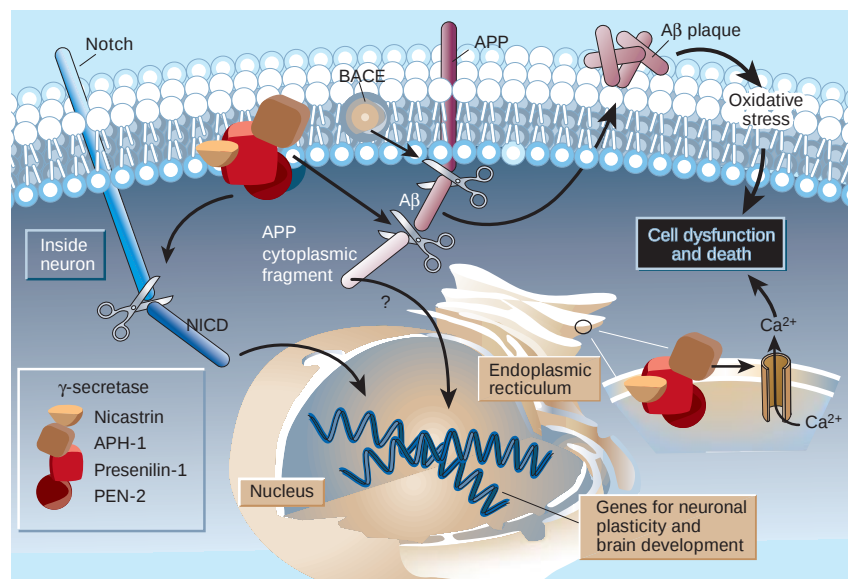
was then discovered when researchers knocked out the presenilin-1 gene in mice (reviewed in ref. 2). The animals died as embryos, and had severe defects in brain development that were indistinguishable from the defects in mice lacking Notch. This is because presenilin-1 is required not only to cleave APP and generate  $A\beta$ , but also to cleave Notch after Notch has detected and bound a partner protein. An intracellular fragment of Notch is then released, and regulates gene expression in the neuronal nucleus. It has been suggested<sup>4</sup> that an intracellular fragment of APP, generated by  $\gamma$ -secretase, likewise moves to the nucleus and regulates gene expression.

But it soon became clear that presenilin-1 cannot work alone to cleave APP and Notch, and a search began for other proteins that might be involved. APP and Notch have been highly conserved during evolution, which not only attests to their physiological importance, but also means that molecular-genetic analyses of fruitflies and worms can be used to investigate their cleavage. Such studies have found that four proteins seem to contribute to  $\gamma$ -secretase activity; these are presenilin-1, nicastrin, A $\phi$ -1 and PEN-2 (Fig. 1, overleaf)<sup>5–7</sup>. It has just been shown that  $\gamma$ -secretase activity can be fully reconstituted with only these four proteins<sup>8</sup>.

But what exactly do these proteins do? To begin to understand this, Takasugi and co-workers<sup>3</sup> first generated fruitfly cells that expressed different combinations of fruitfly nicastrin, A $\phi$ -1 and PEN-2 and determined the effects on cleavage of presenilin-1 (this event having been previously associated with  $\gamma$ -secretase activity). They found that overexpression of A $\phi$ -1 — or A $\phi$ -1 plus nicastrin — stabilized the four-protein complex and simultaneously reduced presenilin-1 cleavage, suggesting that A $\phi$ -1 inhibits the ability of  $\gamma$ -secretase to cleave any of its target proteins. They then showed that, indeed, A $\phi$ -1 reduces the  $\gamma$ -secretase cleavage of APP as well.

To determine the role of PEN-2 in the  $\gamma$ -secretase quartet, the authors used RNA interference to target and degrade the messenger RNA encoding PEN-2, thereby reducing production of the protein, in fruitfly cells, mouse and human brain neurons, and human tumour cells. This resulted in decreased  $\gamma$ -secretase activity. Further experiments in which a fragment of APP was added confirmed that A $\phi$ -1 inhibits, whereas PEN-2 promotes, the production of  $A\beta$ . These findings advance our understanding of an enzyme activity that is important in both brain development and Alzheimer's disease, and identify new protein targets for drugs to prevent or treat this disorder. But the results also raise new questions, and reveal further hurdles to treating Alzheimer's disease.

One general question is whether the



**Figure 1** The  $\gamma$ -secretase protein quartet, and its roles in brain development and Alzheimer's disease. Presenilin-1, nicastrin, APh-1 and PEN-2 form a functional  $\gamma$ -secretase complex, located in the plasma membrane and endoplasmic reticulum (ER) of neurons. The complex cleaves Notch (left) to generate a fragment (NICD) that moves to the nucleus and regulates the expression of genes involved in brain development and adult neuronal plasticity. The complex also helps in generating the amyloid  $\beta$ -peptide ( $A\beta$ ; centre). This involves an initial cleavage of the amyloid precursor protein (APP) by an enzyme called BACE (or  $\beta$ -secretase). The  $\gamma$ -secretase then liberates  $A\beta$ , as well as an APP cytoplasmic fragment, which may move to the nucleus and regulate gene expression. Mutations in presenilin-1 that cause early-onset Alzheimer's disease enhance  $\gamma$ -secretase activity and  $A\beta$  production, and also perturb the ER calcium balance. Consequent neuronal degeneration may result from membrane-associated oxidative stress, induced by aggregating forms of  $A\beta$  (which create  $A\beta$  plaques), and by the perturbed calcium balance.

functions of APh-1 and PEN-2 are the same in Notch cleavage as in APP cleavage. This is important not only for our understanding of Notch signalling, but also from a clinical perspective: existing therapeutic  $\gamma$ -secretase inhibitors decrease  $A\beta$  production, yet probably have serious side effects because they also inhibit Notch cleavage<sup>6</sup> (which is important in adults as well as embryos). In addition, although various environmental signals — such as growth-factor, neurotransmitter and cytokine molecules — affect the expression and processing of APP and Notch<sup>9,10</sup>, we do not yet know if and how such signals affect  $\gamma$ -secretase activity. Another question is whether the expression of the four proteins varies during development, although this seems likely, given the obvious importance of  $\gamma$ -secretase during brain development. And are any of the proteins regulated at a level beyond the expression of their genes, for example by the covalent attachment of phosphate groups?

More specific questions concern how the members of the  $\gamma$ -secretase quartet interact, both physically and functionally, with one another and with their substrates. Physically, they seem to bind to one another; quite how is unknown, but they probably contain protein–protein interaction domains similar to those of other membrane proteins, such as subunits of ion channels. Functionally,

presenilin-1 is believed to be the enzyme that actually cleaves APP and Notch<sup>2</sup>. If so, then nicastrin, APh-1 and PEN-2 might regulate  $\gamma$ -secretase activity by modifying either presenilin-1's enzymatic activity or its association with substrates. Finally, the  $\gamma$ -secretase complex presumably functions in the plasma membrane, where Notch and

APP reside. But considerable evidence suggests that presenilin-1 occurs primarily in the endoplasmic reticulum, a network of internal membranes, so this presents another puzzle.

Returning to Alzheimer's disease, one burning question is how presenilin-1, nicastrin, APh-1 and PEN-2 contribute to the aberrant APP processing and neuronal degeneration seen in this disorder. Besides increasing  $A\beta$  production, presenilin-1 mutations lead to a large increase in the number of calcium ions in the endoplasmic reticulum; this may contribute to defects in neuronal communication and neuronal death<sup>11</sup>. It will be interesting to see if and how these two consequences of presenilin-1 mutations are linked. It is not impossible that the alterations in  $\gamma$ -secretase activity and APP processing caused by presenilin mutations and other genetic and environmental factors are secondary to a primary disturbance in calcium regulation or oxidative stress. It will be important to find out whether nicastrin, APh-1 or PEN-2 modifies the effects of presenilin-1 mutations on neuronal calcium balance, and on the vulnerability of neurons.

Mark P. Mattison is in the Laboratory of Neurosciences, National Institute on Aging, Gerontology Research Center, 5600 Nathan Shock Drive, Baltimore, Maryland 21224, USA.  
e-mail: mattisonm@grc.nia.nih.gov

1. Haass, C. & De Strooper, B. *Science* **286**, 916–919 (1999).
2. Selkoe, D. J. *Curr. Opin. Neurobiol.* **10**, 50–57 (2000).
3. Takasugi, N. et al. *Nature* **422**, 438–441 (2003).
4. Leissring, M. A. et al. *Proc. Natl Acad. Sci. USA* **99**, 4697–4702 (2002).
5. Yu, G. et al. *Nature* **407**, 48–54 (2000).
6. Francis, R. et al. *Dev. Cell* **3**, 85–97 (2002).
7. Behr, D. & Shearman, M. S. *Biochem. Soc. Trans.* **30**, 534–537 (2002).
8. Edbauer, D. et al. *Nature Cell Biol.* (in the press).
9. Baron, M. et al. *Mol. Membr. Biol.* **19**, 27–38 (2002).
10. Panegyres, P. K. *Rev. Neurosci.* **12**, 1–39 (2001).
11. Chan, S. L. et al. *Neuromol. Med.* **2**, 167–196 (2002).

## Quantum computing

# Logic gateway

Andrew Steane

Two groups have created logic gates using pairs of trapped ions. As components of a quantum computer, these gates have the potential to form part of a scaled-up, workable system.

Quantum computing often features in these pages, owing to the lively pace of research in this area. The main aim is to build a large working quantum computer, but the path towards that goal promises to reveal some wonderful physics; subtle and surprising effects are confidently expected, such as quantum error correction. Also, the rich nature and behaviour of quantum entanglement — the special correlation that causes composite quantum systems to defy description in terms of their constituent parts

— is not fully understood, even for systems of only a few particles, and experimentation is essential to understand larger systems. The techniques and language of quantum computing are very powerful. They will almost certainly give birth to a new generation of ultra-precise experimental methods based on the interference properties of entangled systems, whether or not efficient quantum computers are eventually realized.

Two papers in this issue<sup>1,2</sup> (on pages 408 and 412) report notable progress in what is



the essence of quantum computing, namely the controlled manipulation of entanglement. Both experiments involve quantum control of charged atoms confined in a trap in a vacuum and manipulated by laser pulses. They realize, by different methods, the basic computing operation — a controlled logic gate between a pair of quantum bits (qubits). It should be noted, however, that the other main ingredients for quantum computing — precise state-preparation and readout, and high-quality quantum memory — are already reliable and taken for granted in such ion-trap technology. The theme here is not the mere demonstration of entanglement, but something more important which I will bring out after discussing the experiments themselves.

Leibfried and co-workers<sup>2</sup> have invented and demonstrated a new technique to achieve a robust, two-qubit logic gate using trapped ions. A pair of beryllium ions sits in a single trap, and the hyperfine structure of their electronic energy states enables them to store extremely stable qubits. These internal qubit states can be represented by  $|\uparrow\rangle$  and  $|\downarrow\rangle$ . The method is to apply an oscillating force to the trapped ion pair, but one that acts only when the ions are in different internal states: when the joint state is either  $|\uparrow\downarrow\rangle$  or  $|\downarrow\uparrow\rangle$ , the forces on the ions are not balanced and they oscillate slightly towards and away from each other. This changes the average repulsive Coulomb force between the charged ions, and hence, while the force is acting, the energy of the states  $|\uparrow\downarrow\rangle$  or  $|\downarrow\uparrow\rangle$  is different from half the energy of the state  $|\uparrow\uparrow\rangle$ . This is the so-called phase gate, which is logically equivalent to the 'controlled NOT' operation of conventional electronics.

The oscillating force is generated by a pair of counter-propagating beams, directed at the ions. A frequency difference of 6.126 MHz between the lasers sets up a highly accurate oscillation frequency for the force. This frequency difference is deliberately chosen to be close to the natural 6.1-MHz vibration frequency of the ions in the trap, taking up an idea that Milburn *et al.*<sup>3</sup> first proposed. Under these conditions, the quantum state of the ion pair is translated around a loop in position–momentum space, as the relative phase between the ion and laser oscillations changes. After 39  $\mu$ s, the ion and laser oscillations come back into phase with each other, closing the loop traced in position–momentum space. When the ions are prepared in a superposition of  $|\uparrow\rangle$  and  $|\downarrow\rangle$  states (which can be done using standard laser pulses in which the ions evolve independently), the net effect is to entangle them, because the final quantum phase of each depends on the other.

The quantum logic operation is determined purely by the area of the loop traced in position–momentum space, and not by

its shape. As a result, the gate is robust — the motion of the ions does not have to be controlled in detail (they can have thermal motion, for example) and the size of the force (determined by the laser intensity) need not be stable throughout. Leibfried *et al.*<sup>2</sup> report the controlled creation of a maximally entangled state with a success rate of 97%; most of the 3% error is due to a spontaneous energy-emission process, which is known to be avoidable in other ion species.

Schmidt-Kaler *et al.*<sup>1</sup> have realized a logic gate between two trapped ions with the added ingredient that the ions can be addressed individually by focused laser beams. Their logic-gate method is based on the ideas of Cirac and Zoller<sup>4</sup>, in which vibrational motion of the ion pair is excited selectively by imparting momentum to the ions when they absorb, or are stimulated to emit, single photons. In earlier experiments, the small separations of the trapped ions meant that they could not be controlled and read out individually. In this experiment<sup>1</sup>, two calcium ions, separated by 5.3  $\mu$ m, are addressed by a laser beam of width 2.5  $\mu$ m. Reliable discrimination of all four double-qubit states ( $|\uparrow\uparrow\rangle$ ,  $|\downarrow\downarrow\rangle$ ,  $|\uparrow\downarrow\rangle$  and  $|\downarrow\uparrow\rangle$ ) in the readout is possible. This team has overcome a number of experimental difficulties using various techniques, among them an adaptation of the composite-pulse methods developed for nuclear magnetic resonance technology: where any laser pulse creates an

unwanted evolution of part of the joint state, the team identifies sequences of pulses that undo that evolution but still produce the desired (non-zero) net effect.

Until now, entangled trapped ions had been created in only one laboratory — that of David Wineland at the National Institute of Standards and Technology in Colorado, USA (the group are the authors of Leibfried *et al.*<sup>2</sup>). It is a sign of growing experimental maturity that the essential experiments are now being done successfully in more than one laboratory. Furthermore, there is an important theme running through both of these papers<sup>1,2</sup>: we are seeing here efforts not merely to attain some sort of entanglement manipulation in ion traps, but to do it well.

This is essential if quantum computing is to be achieved. A quantum computer will rely absolutely on qubits and logic gates that have excellent precision and can be taken for granted. The experiments by Leibfried *et al.*<sup>2</sup> and Schmidt-Kaler *et al.*<sup>1</sup> represent, for me, the first hint that there is a serious possibility of making logic gates, precise to one part in a thousand or even ten thousand, in a system that could be scaled up to many qubits. ■

Andrew Steane is at the Centre for Quantum Computation, Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK.  
e-mail: a.steane1@physics.ox.ac.uk

1. Schmidt-Kaler, F. *et al.* *Nature* **422**, 408–411 (2003).

2. Leibfried, D. *et al.* *Nature* **422**, 412–415 (2003).

3. Milburn, G. *et al.* *Fortschr. Phys.* **48**, 801–810 (2000).

4. Cirac, J. I. & Zoller, P. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).

## Palaeontology

# Combing the primate record

Robert D. Martin

Fossils of bushbabies and lorises reported from deposits of the Fayum Depression in Egypt extend the known record for this group of primates from 20 million years to approximately 40 million years ago.

Primate evolution attracts special interest because of its direct relevance to human origins. The fossil record remains extremely sparse, however, and that is particularly true of a group called the strepsirrhines, which comprises the bushbabies and lorises — together, the lorisiforms — and the lemurs. Part of that picture now changes radically with the report, by Seiffert, Simons and Artia on page 421 of this issue<sup>1</sup>, of plausible early members of the bushbaby and loris lineages. The strepsirrhines are the 'sister group' of the haplorhines — the tarsiers and higher primates, which include ourselves. The relationships among the main groups of living primates are shown in Fig. 1.

The lemurs are found only in Madagascar, and consist of 15 modern genera and 54 species<sup>2</sup>. Their palaeontological record is virtually zero. The record of lorisiforms, a

smaller assemblage of 8 modern genera and 32 species<sup>2</sup>, was scarcely better: until now, the earliest known fossil representatives were three early Miocene genera from Kenya (*Komba*, *Mioeuoticus* and *Progalago*), up to 20 million years old<sup>3</sup>. Seiffert and colleagues' discovery of isolated teeth and jaw fragments of two new genera, *Karanisia* and *Saharagalago*, from late middle Eocene deposits of Egypt, improves matters considerably. Both forms share unique dental features with modern lorisiforms, notably a prominent fourth cusp (hypocone) on each upper molar, contributing to a distinctive concave posterior tooth margin. In one fell swoop, *Karanisia* and *Saharagalago*, dated at between 37 and 41 million years old, have doubled the age of the known fossil record for lorisiform primates and illuminate the African diversification of lorisiforms.

An overall assessment of morphological,

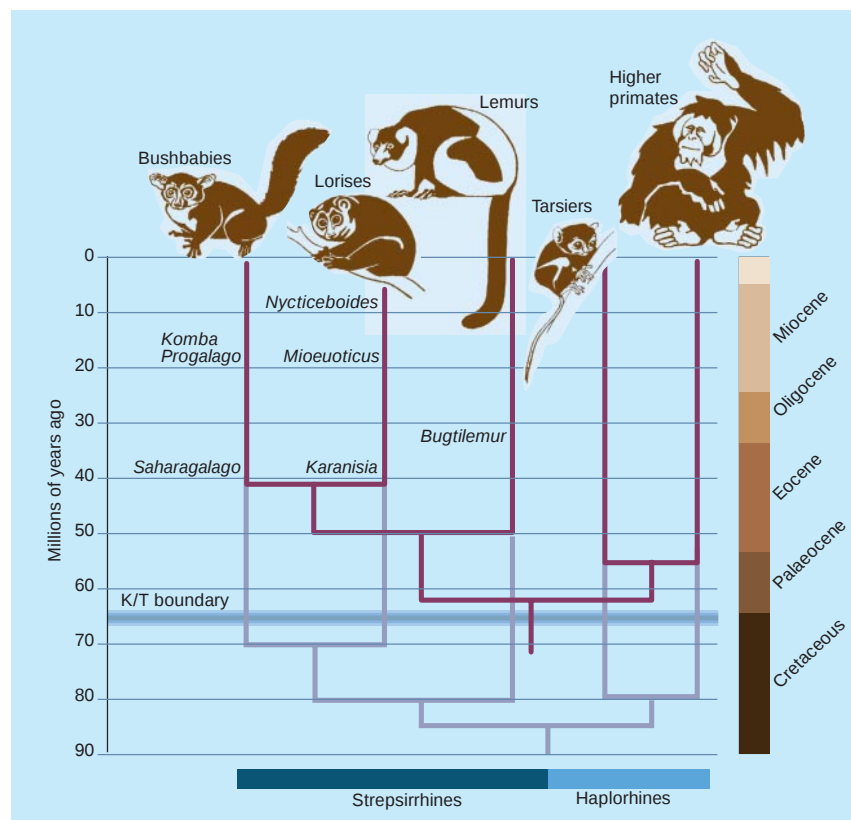


Figure 1 A simplified primate evolutionary tree. The two main groups of living primates, the strepsirrhines and haplorhines, are shown, and the new fossil forms (*Saharagalago* and *Karanisia*) described by Seiffert and colleagues<sup>1</sup> are included. These specimens double the known age of the lineages leading to bushbabies and lorises (together, the lorisiforms), and increase the number of known lorisiform genera by half. The previously known genera are *Komba*, *Progalago* and *Mioeoticus* from the early Miocene of East Africa, and *Nycticeboides* from the late Miocene of Pakistan. The only known fossil form potentially related to the other strepsirrhine group, the lemurs, is *Bugtilemur* from the early Oligocene of Pakistan. A direct reading of the fossil record indicates that primates began to diversify just after the Cretaceous/Tertiary (K/T) boundary. But, given the incompleteness of the fossil record, diversification may have occurred at least 20 million years before the K/T boundary, as indicated by the shadow tree. (Primate icons drawn by Lucrezia Beerli-Bieler.)

chromosomal and molecular evidence supports the view that the strepsirrhines and haplorhines diverged early in primate evolution<sup>4</sup>. Morphologically, strepsirrhines are clearly distinguished by their retention of various primitive features, such as the rhinarium (a naked area of moist skin surrounding the nostrils), and by a uniquely derived toothcomb. This structure is used both for feeding (for instance, scooping pulp from fruit and scraping gum from trees) and for grooming the fur, and is formed by the canines and incisors in the lower jaw (six teeth in all). These teeth are elongated and bilaterally flattened into tines, and their narrow crowns angle sharply forwards from the roots, almost horizontally. Hairs dragged between these tooth crowns during grooming leave conspicuous grooves, clearly recognizable on fossilized teeth<sup>5</sup>. A convincing indication that *Karanisia* is, indeed, a strepsirrhine is that a lower-jaw fragment shows typically compressed tooth sockets for the narrow roots of the canine and incisors, and

the crown of an isolated canine shows tiny lateral grooves worn by the passage of hair. This neatly confirms the prediction that the toothcomb was present in early strepsirrhines, and also provides direct fossil evidence of the occurrence of grooming behaviour 40 million years ago.

Soon after the initial split between the toothcombed strepsirrhines and the haplorhines, there was a divergence between the lemurs and the Afro-Asian lorisiforms<sup>4,6,7</sup>. The lorisiform lineage then divided into bushbabies and lorises (Fig. 2, overleaf), a subdivision that may also be documented by the new fossils. Molars of *Saharagalago* strikingly resemble those of modern bushbabies, whereas those of *Karanisia* are generally similar to loris molars, particularly those of the African angwantibo (*Arctocebus*). However, one of the most striking differences between modern bushbabies and lorises involves locomotor adaptations: lorises typically climb sluggishly (reflecting their very low basal metabolic rates), whereas



#### 100 YEARS AGO

The investigation of the properties of radium salts has led to many remarkable results, among which those contributed by MM. P. Curie and A. Laborde to the current number of the *Comptes rendus* are not the least remarkable. They adduce evidence to show that radium salts give off heat continuously. ... Two small bulbs, one containing 1 gram of a radiferous barium chloride containing about 1/6 of its weight of radium chloride and the other containing a similar weight of ordinary barium chloride, were placed under similar thermal conditions with a junction of a thermocouple in each bulb. The bulb containing the radium preparation proved to be 1°·5 hotter than the other, and this temperature difference was maintained. An independent confirmation was obtained with the Bunsen ice calorimeter. At the moment the radium bulb was introduced, the mercury, which was previously stationary, commenced to move along the tube with a perfectly uniform velocity, and on the bulb being taken out the mercury stopped. ... the authors conclude that a gram of pure radium would give off a quantity of heat of the order of 100 calories per hour, or 22,500 per gram-atom per hour, a number comparable with the heat of combustion in oxygen of a gram-atom of hydrogen. The disengagement of such a quantity of heat cannot be explained by the assumption of any ordinary chemical transformation, and this excludes the theory of a continuous modification of the atom. The heat evolution can only be explained by supposing that the radium utilises an external energy of unknown nature.

From *Nature* 26 March 1903.

#### 50 YEARS AGO

The Cecil Peace Prize of the Association of Universities of the British Commonwealth has been awarded for 1952 to K. H. Dawson, of the London School of Economics (at present at the Graduate School, Princeton University). The subject for the competition for 1953 is "Has the United Nations Organization been successful in carrying out the objects for which it was formed as defined in the Preamble and Chapter 1 of the Charter? Do you recommend any, and if so, what amendments in the terms of the Charter?" Essays must be sent before November 1 to the Secretary, Association of Universities of the British Commonwealth, 5 Gordon Square, London, W.C.1, from whom further details can be obtained.

From *Nature* 28 March 1953.



Figure 2 Lorisiforms here and now — a bushbaby (left) and a loris.

bushbabies commonly leap actively and show conspicuous elongation of the tarsal bones of the hindlimb. Isolated tarsals from *Progalago* show moderate elongation, but in the absence of limb-bone fossils of *Saharagalago* or *Karanisia*, we cannot say whether these earlier forms already showed the expected divergence in locomotion. Modern bushbabies also clearly differ from lorises in possessing molar-like posterior premolars. The discovery of similar teeth in *Saharagalago* would provide a valuable additional test of the inferred relationship to modern bushbabies.

*Saharagalago* and *Karanisia* are the latest discoveries in four decades of fossil-hunting by Elwyn Simons and colleagues in the Fayum Depression. The Fayum site, on the eastern margin of the Sahara in Egypt, contains extensive sediments spanning the late Eocene and the early Oligocene. It provides the best known record for the early evolution of modern mammals in Africa. This single site has yielded a remarkable diversity of primate fossils, most being of higher primates, although previous finds include *Afrotarsius* (possibly related to modern tarsiers) and *Plesiopithecus* (an aberrant early strepsirrhine offshoot).

More broadly, Seiffert and colleagues<sup>1</sup> conclude that the new finds are compatible with a date of 50–53 million years ago for the last common ancestor of strepsirrhines, with Afro-Arabia being the location of that ancestor. But this conclusion stems from a direct reading of an inadequate fossil record. We still have no fossil primates from sub-Saharan Africa before the Miocene, and the fossil record for Madagascar lemurs remains nil. My own alternative interpretation stems from statistical considerations<sup>8</sup> indicating that such gaps in the primate fossil record have led to a serious underestimation of divergence times. Furthermore, molecular phylogenies for placental mammals<sup>9</sup> have revealed a group of endemic African mammals (Afrotheria) that does not include the primates, suggesting that primates originated elsewhere. Various molecular trees also indicate that primates originated much earlier than generally accepted, about 90 million years ago. One possibility is that strepsirrhines originally inhabited Indo-Madagascar,

rather than Africa, and that lemurs became isolated when Madagascar separated from India. Subsequently, lorises could have migrated to Africa after India collided with Asia, reaching Africa during the Eocene.

Interestingly, the only known fossil primate with direct affinities to lemurs (*Bugtilemur*; Fig. 1) was reported from Oligocene deposits in Pakistan<sup>10</sup>: its dentition closely resembles that of modern dwarf lemurs. *Bugtilemur* raises more questions than it answers, however, and shows that we still have much to learn about the early evolution of our primate relatives.

Robert D. Martin is in Academic Affairs at The Field Museum, 1400 S. Lake Shore Drive, Chicago,

Illinois 60605-2496, USA.

e-mail: rdmartin@fieldmuseum.org

1. Seiffert, E. R., Simons, E. L. & Attia, Y. *Nature* **422**, 421–424 (2003).
2. Groves, C. P. *Primate Taxonomy* (Smithsonian Institution Press, Washington DC, 2001).
3. Phillips, E. M. & Walker, A. in *The Primate Fossil Record* (ed. Hartwig, W. C.) 83–95 (Cambridge Univ. Press, 2002).
4. Martin, R. D. *Primate Origins and Evolution: A Phylogenetic Reconstruction* (Princeton Univ. Press, 1990).
5. Rose, K. D., Walker, A. C. & Jacobs, L. L. *Nature* **289**, 583–585 (1981).
6. Yoder, A. D. *Evol. Anthropol.* **6**, 11–22 (1997).
7. Yoder, A. D., Cartmill, M., Ruvolo, M., Smith, K. & Vilgalys, R. *Proc. Natl Acad. Sci. USA* **93**, 5122–5126 (1996).
8. Tavaré, S., Marshall, C. R., Will, O., Soligo, C. & Martin, R. D. *Nature* **416**, 726–729 (2002).
9. Murphy, W. J. et al. *Science* **294**, 2348–2351 (2001).
10. Marivaux, L. et al. *Science* **294**, 587–591 (2001).

### Quantum physics

## Wheels within wheels

Jurgen H. Smet

Quasi-particles, an ingenious dodge used to simplify calculations on vast systems of interacting particles, seem to account for the fractional quantum Hall effect. But do we now need a further generation of quasi-particles?

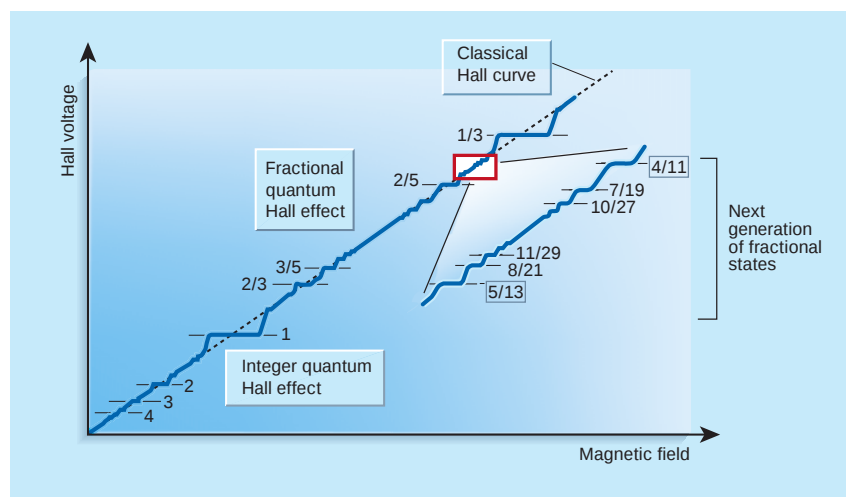
Problems involving many strongly interacting bodies are pervasive in physics. For instance, in a cubic centimetre of condensed matter, typically  $10^{23}$  electrons repel one another and interact with a comparable number of positively charged nuclei. The motion of one electron elicits a response from all of the others. So it is no wonder that the search for an exact solution is usually a desperate undertaking. Ingenious recipes have been developed, based on quantum field theory, to tackle these problems: they identify fictitious entities — quasi-particles — that recast the system of strongly interacting real bodies into a simpler one, composed of weakly or even non-interacting bodies, while still capturing the essential physics. Landau's quasi-electrons — bare electrons dressed with a cloud of positive charge — are a celebrated example that successfully describes the behaviour of metals. More recently, composite fermions<sup>1</sup> — electrons in a different guise — have emerged to account in single-particle terms for the fractional quantum Hall effect<sup>2</sup>, an electron–electron correlation phenomenon *par excellence*. Pan et al.<sup>3</sup> now report in *Physical Review Letters* that the story goes on: apparently, residual interactions among these composite fermions produce a second generation of composite fermions.

Quantum Hall effects<sup>4,5</sup> arise when electrons constrained to move in a plane are exposed to a perpendicular magnetic field. They are quantum mechanical descendants of a classical effect discovered by Edwin Hall more than a century ago. He observed that a current-carrying conductor in the presence of a field develops a voltage that is perpendicular

to both the current flow and the field. Ever since, a measurement of the Hall voltage has been a valuable characterization method in solid-state physics, because it reveals the number as well as the sign of current-carrying charges. Its application to clean, near-perfect two-dimensional conductors at low temperatures brought a new twist. Here, the Hall voltage does not simply rise linearly with the applied field. Instead, it shows plateaux, as if the Hall voltage is frozen near specific field values (Fig. 1, overleaf). Across the plateaux, the voltage drop in the direction of the current flow vanishes; this is the second hallmark of the quantum Hall effects.

The magnetic field sends the electrons into circular orbits. In classical physics, any radius is allowed. Quantum mechanics, however, dictates discrete values for the radius, much as it imposes distinct Bohr orbits on an atom. This is an outcome of the discrete character of magnetic flux: the applied field derives from many flux quanta, each contributing the smallest unit of magnetic flux to the total. According to the laws of quantum mechanics, only electron orbits that enclose exactly one such quantum or multiple quanta of magnetic flux are legitimate. Like the Bohr orbits, each of these orbits has a discrete energy associated with it — a Landau level. At fixed field, the larger the radius of an orbit, the higher its energy. The electrons are distributed among the orbits or Landau levels so as to minimize the total energy, keeping in mind that each orbit fits only one electron. However, many orbits of equal size (and thus energy) are spread throughout the sample. To be precise, each Landau level can accommodate as many





**Figure 1 The Hall effect.** The Hall voltage develops when driving a current through a conductor exposed to a perpendicular magnetic field. In classical physics, it follows a straight line. But quantum mechanics forces the electrons to occupy discrete energy levels. Whenever an integral number of these levels is filled, a plateau appears in the Hall voltage. At higher fields, when the lowest level is only partially filled, additional plateaux arise by virtue of interactions among the electrons — the fractional quantum Hall effect. This is equivalent to the integer Hall effect for ‘quasi-particles’ made up of an electron plus two flux quanta. But it can be taken further: add two more flux quanta to the composite quasi-particle, and you might expect new plateaux at, for instance, the fractions shown in the enlargement (schematic only; the plateaux are yet to be confirmed). Pan *et al.*<sup>3</sup> have seen signs of some of these new fractional states (in boxes). With increasing sample quality the number of plateaux seems to grow, such that the Hall curve starts to show fractal characteristics.

electrons as the total number of flux quanta that thread the sample. As the field is raised and the number of flux quanta increases, Landau levels can take up ever more electrons, and higher energy levels are successively depopulated. The filling factor  $\nu$  denotes the number of filled levels. When  $\nu$  takes on an integer value, the system will resist the addition of an extra electron, as it must broach a new Landau level with higher energy. This ‘incompressibility’ is at the heart of the integer quantum Hall effect<sup>4</sup>.

Its cousin, the fractional quantum Hall effect<sup>5</sup>, ensues mainly at higher fields when one Landau level is occupied and the filling takes on a fraction that can be expressed as a ratio of integers,  $\nu = p/q$ , with  $p$  and  $q$  integers. Despite its experimental resemblance, it cannot be accounted for directly in the above picture, considering only the motion of a single electron. If the one occupied level is only partially filled, why the incompressibility? Early on, it was recognized that all electrons must participate to bring about this effect<sup>6</sup>. At many of these fractional fillings, electrons apparently succeed in becoming arranged within the Landau level so as to significantly reduce their mutual repulsion.

Much later, it was realized that there is no need to track all electrons to understand this phenomenon. At high fields, compound particles come on the scene, each assembled from an electron and two flux quanta<sup>7</sup> (or more generally, an even number of them). This bond between electrons and flux quanta turns out to be a natural way for electrons to

avoid one another, and the resulting quasi-particles, named composite fermions, may for many purposes be viewed as non-interacting. They, too, are forced by a field into circular orbits, which must obey the laws of quantum mechanics<sup>8</sup>. But, unlike electrons, they experience only an effective field, greatly reduced from the applied field by an amount equal to the field produced by all the flux quanta of their fellow composite fermions. The discrete orbits again have associated Landau levels. Filling these gives the integer quantum Hall effect for composite fermions. Sure enough, it occurs at precisely those applied fields where the fractional quantum Hall effect is routinely observed.

Now Pan *et al.*<sup>3</sup> have demonstrated experimentally that this can be taken still further. A partially filled composite fermion level may in turn bring forth a new generation of composite fermions. At fillings between  $1/3$  and  $2/5$ , two Landau levels of two-flux-quanta composite fermions are filled, one completely, the other partially. Pan *et al.* observed precursors of the fractional quantum Hall effect associated with those composite fermions in the partially filled level. This can be understood if composite fermions collect an extra pair of flux quanta to form four-flux-quanta composite fermions instead, which — as one might guess by now — undergo the integer quantum Hall effect. The other two-flux-quanta composite fermions in the completely filled level are left intact and cohabit with the higher-order composite fermions.

Sample imperfections and the quantum Hall effect suffer a love-hate relationship. Disorder is essential to bring out wide plateaux in the Hall voltage, but too much will overwhelm and destroy more fragile quantum Hall states. The new fractional states, unveiled by further reducing sample imperfections, are in strength and appearance a natural progression from previously reported states. As crystal growers achieve ever lower levels of disorder, and laboratories reach ever lower electron temperatures, we may expect to see additional fractional states. The data of Pan *et al.*<sup>3</sup> support an iterative scheme, where composite fermions in a partially filled level accumulate a growing number of flux quanta, as an intuitive algorithm to predict the next sequence of states to be discovered. This straightforward picture is enticing, and works marvellously for the fractional quantum Hall states discerned so far. But it has no rigorous foundation. Indeed, preceding theoretical work outside<sup>9,10</sup> and within<sup>11</sup> the framework of composite fermions suggests that some of the newly discovered states are unstable. There is clearly a need to reconsider these issues and to examine the residual interactions that might prompt the mutation into higher-order composites.

The recurring pattern of transforming fractional into integer quantum Hall states also bears on the geometrical properties of the Hall curve itself. As pointed out some time ago<sup>12</sup>, the curve shows self-similarity, as it unfolds like a fractal<sup>13</sup>, ever more detail being added as sample quality improves. It seems that the integer quantum Hall curve has been applied as a ‘fractal generator’ or template<sup>12</sup>, the iterative flux attachment translating into the sequential replacement of segments with transformed copies of the template (Fig. 1). The work by Pan *et al.* urges us to pay more attention to this fractal-like structure: it may guide us to a deeper understanding of the quantum Hall effects. ■

Jurgen H. Smet is at the Max-Planck-Institute for Solid State Physics, Heisenbergstrasse 1, D-70569 Stuttgart, Germany.  
e-mail: j.smet@fkf.mpg.de

- Heinonen, O. (ed.) *Composite Fermions: A Unified View of the Quantum Hall Regime* (World Scientific, Singapore, 1998).
- Das Sarma, S. & Pinczuk, A. (eds) *Perspectives in Quantum Hall Effects: Novel Quantum Liquids in Low-Dimensional Semiconductor Structures* (Wiley, New York, 1997).
- Pan, W. *et al.* *Phys. Rev. Lett.* **90**, 016801 (2003).
- von Klitzing, K., Dorda, G. & Pepper, M. *Phys. Rev. Lett.* **45**, 494–497 (1980).
- Tsui, D. C., Stormer, H. L. & Gossard, A. C. *Phys. Rev. Lett.* **48**, 1559–1562 (1982).
- Laughlin, R. B. *Phys. Rev. Lett.* **50**, 1395–1398 (1983).
- Jain, J. K. *Phys. Rev. Lett.* **63**, 199–202 (1989).
- Halperin, B. I., Lee, P. A. & Read, N. *Phys. Rev. B* **47**, 7312–7343 (1993).
- Haldane, F. D. M. *Phys. Rev. Lett.* **74**, 2090–2093 (1995).
- Wójs, A. & Quinn, J. J. *Phys. Rev. B* **61**, 2846–2854 (2000).
- Mandal, S. S. & Jain, J. K. *Phys. Rev. B* **66**, 155302 (2002).
- Mani, R. G. & von Klitzing, K. Z. *Phys. B* **100**, 635–642 (1996).
- Mandelbrot, B. B. *The Fractal Geometry of Nature* (Freeman, New York, 1979).

## Evolutionary biology

## Asexual means of mutation control

*Proc. R. Soc. Lond. B* doi:10.1098/rspb.2002.2314 (2003)

The crustacean *Darwinula stevensoni* has managed to reproduce without sex for at least 20 million years, making it the oldest known asexual animal species. Isabelle Schön and Koen Martens now report that it is also remarkable for the fact that different individuals and populations vary very little from one another genetically.

This is actually not what one would expect. As the separate lineages of asexual species go through the generations, mutations should accumulate, creating differences in the genome that will not be removed during sexual reproduction: one explanation for the prevalence of sex in the natural world is that, by shuffling DNA, it helps to get rid of mutations, which will mostly be damaging.

Yet Schön and Martens found that the genetic distance between individuals of *D. stevensoni* is only about a quarter of that between members of a closely related sexual species. They speculate that *Darwinula* has special mechanisms to prevent it from accumulating mutations, such as exceptionally good DNA repair. **John Whitfield**

## Cell biology

## Muscle mice

*J. Cell Biol.* **160**, 909–918 (2003)

**Cell-based treatments for muscle-wasting disorders such as muscular dystrophy have tended to focus on replacing like with like. But adult muscle cells are hard to culture, and age rapidly. In contrast, less specialized cell types such as stem cells are easier to multiply up in culture, and may contribute more fully to skeletal-muscle regeneration.**

Cosimo De Bari and colleagues have now injected bone-marrow-derived adult stem cells, taken from human knee joints, into immunodeficient mice with a form of toxin-induced muscle wasting. The multi-talented cells homed in on the damaged area, producing functional satellite cells (muscle-cell progenitors) and contributing to muscle fibres. As the stem cells differentiated into these other cell types, they did so naturally — their gene-expression patterns mimicked those seen during normal embryonic muscle formation.

The stem cells also helped to repair damage in a mouse model of Duchenne muscular dystrophy. Muscle fibres once again made dystrophin, a structural protein whose loss is implicated in wasting diseases, and more of mechano growth factor, involved in maintaining and repairing skeletal muscle. The findings suggest that

the transplant of adult stem cells, derived from patient biopsies, might be an effective treatment for Duchenne muscular dystrophy.

**Helen R. Pilcher**

## Earth science

## Stream flow and shaken soil

*Geophys. Res. Lett.* doi:10.1029/2002GL016618 (2003)

Shortly after an earthquake, the flow of streams in the vicinity often increases. This phenomenon has baffled geologists for many years. Michael Manga and colleagues now propose that shaking forces water out of surface soils and produces the extra stream input.

Previous explanations for this were of two types. First, that compression following earthquakes squeezes water-bearing rocks, forcing more water out of them. Second, that the earthquake-induced strain on rocks creates fissures or pores that allow water to percolate out faster. From their studies of the flow history of Sespe Creek in southern California, Manga *et al.* come to a different conclusion. Sespe Creek has experienced 48 strong earthquakes since its monitoring began in 1928. Not all of those events produced increases in water flow, but when they did, it was regardless of whether the watershed was compressed or expanded.

Manga *et al.* hypothesize that seismic shaking compacts surface soils and forces more water out of pores between soil particles. Laboratory tests on a shaking table suggest that sufficient volumes, typically equivalent to a few millimetres of extra rainfall, can be produced from soils by earthquakes.

**Tom Clarke**

## Neurobiology

## Spine conducts hip-hop

*Science* **299**, 1889–1892 (2003)

**Most mice scamper — but without certain genes, they hop like rabbits. Klas Kullander *et al.* now show that these genes wire up a neuronal circuit in the spinal cord that may trigger walking.**

This spinal-cord circuit, called the central pattern generator, is thought to stimulate an innate, alternating left–right gait under directions from the brain. In support of this, the authors found that, in normal mice, electrical signals shoot out alternately from the left and right sides of the lower spinal cord. But in mice genetically engineered to lack either ephrinB3 or its receptor EphA4 — molecules that guide growing nerves — some nerves stray across the middle of the spinal cord. Electrical signals shoot out from both sides, triggering a kangaroo-kick.

These molecules are the first to be identified that direct the development of

nerves in the mammalian central pattern generator, and may lead researchers to the equivalent circuits in humans. This could ultimately help patients who have suffered spinal-cord injury, by highlighting which areas to salvage — and, perhaps, how best to stimulate them to recreate a walking movement.

**Helen Pearson**

## Nanoelectronics

## Silicon springs leaks

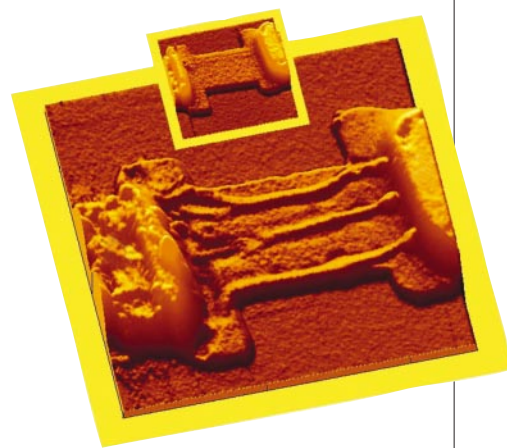
*Appl. Phys. Lett.* **82**, 1727–1729 (2003)

Nanoscale silicon devices already face severe obstacles to their creation, and now N. Clement *et al.* show that their prospects are even worse: the devices may be rapidly eaten away by the current passing through them, becoming riddled with hollow channels between their electrical terminals.

Silicon devices could become as small as a few tens of nanometres within a few years. But this poses potential problems, such as current leakage. The high electric fields created in such small devices may also accelerate mobile electrons to such energies that they act like bullets, colliding with silicon atoms and producing electron–hole pairs, boosting the current through the device.

Clement and colleagues suspect that something of this sort is creating the channels (about 45 nm wide) that open up between the terminals in their transistor-like devices (see picture). These terminals are connected by a 20-nm-thick slab of silicon laid down on silicon dioxide — the ‘silicon-on-insulator’ structure being considered for next-generation miniaturization. The channels, which seem to result from heat-induced diffusion of the silicon ions created by electron impacts, play havoc with the current–voltage relationship of the devices, rendering them useless.

**Philip Ball**



Before and after: atomic-force-microscope images show the damage done to a silicon nanowire subjected to a high voltage (the inset shows the nanowire before the voltage was applied).

# Manual dexterity in Neanderthals

These primitive people may have been as handy with their tools as modern humans are.

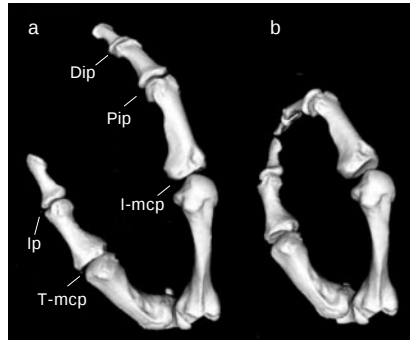
Despite their ability to make and use stone tools, Neanderthals were presumed to have had limited manual dexterity on the basis of the anatomy of their thumb and forefinger<sup>1</sup> — a contention that has been called into question<sup>2–4</sup>. Here we investigate the likely extent of Neanderthal thumb function by using a three-dimensional dynamic simulation that is based on the anatomical details and articular morphology of the thumb and index finger. We find that these digits could make tip-to-tip contact, and conclude that manual dexterity in Neanderthals was probably not significantly different from that of modern humans.

Epoxy casts of the La Ferrassie I Neanderthal thumb and index-finger bones were scanned with a Minolta Vivid 900 laser digitizer to produce three-dimensional polygon mesh models that are accurate to within  $\pm 20$  micrometres (Fig. 1a, b). We used these bone models to generate an articulated computer model with Maya Unlimited software. Movements were simulated by user-assigned ranges of motion at the centre of rotation of each joint with the aid of Maya animation tools.

A saddle-shaped metacarpal-1 base is a key feature for the production of precision grips (that is, the pads of the thumb and the fingers are opposed)<sup>5</sup>. A three-dimensional morphometric analysis indicates that some Neanderthal metacarpal-1 bases approach a condyloid shape because they lack a highly developed palmar beak<sup>4</sup>. The La Ferrassie I metacarpal-1 base has a moderately developed palmar beak that is at the extreme of the modern-human range of variation<sup>4</sup>, making it likely that the range of movements of its trapezium-metacarpal joint is similar to that of modern humans.

In fact, given the open configuration of the Neanderthal trapezium-metacarpal-1 joint, all Neanderthal thumbs were probably more mobile than that of modern humans. But as it has been suggested that Neanderthal thumb movements were restricted<sup>1,6,7</sup>, we minimized the mobility of the thumb by using the middle range of published modern-human flexion/extension values<sup>3,8</sup> and by eliminating metacarpal-1 pronation entirely.

The movement of the index finger is another important factor for producing precision grips. The asymmetry of the second metacarpal head (one half of the index-finger knuckle joint), combined with actions of muscle contraction and joint ligaments, causes the finger to turn towards the thumb in full flexion<sup>9</sup>. Although Neanderthal metacarpal-2 heads are slightly



**Figure 1** The laser-scanned Neanderthal La Ferrassie I thumb and index finger. **a**, **b**, Dorsal views of the thumb (left) and index finger (right) in **a**, their neutral position, and **b**, the fully flexed position. In **b**, the two digits make tip-to-tip contact; the thumb carpometacarpal (bottom) joint is flexed by  $10^\circ$  from its neutral position, the thumb metacarpophalangeal (T-mcp) joint is flexed by  $35^\circ$ , the interphalangeal (Ip) joint is flexed by  $15^\circ$ , the index-finger metacarpophalangeal (I-mcp) joint is flexed by  $45^\circ$ , the proximal interphalangeal (Pip) joint is flexed by  $65^\circ$ , and the distal interphalangeal (Dip) joint is flexed by  $15^\circ$ .

less asymmetrical than those of modern humans<sup>4</sup>, we were again cautious and restricted movement at the metacarpophalangeal joint to flexion/extension only. We assigned published modern-human flexion/extension values to the interphalangeal joints<sup>9</sup> because these joints are functionally equivalent in Neanderthals and modern humans.

Even allowing for significantly limited joint movements relative to modern humans, flexion of the thumb and index finger results in tip-to-tip contact (Fig. 1b). Musgrave<sup>1</sup> suggested that the Neanderthal's short thumb and first-finger proximal

phalanges could have inhibited their precision of movement, but our results indicate that this is not the case. As there is no significant difference between Neanderthals and modern humans in the locations of their muscle and ligamentous attachments<sup>2</sup>, there remains no anatomical argument that precludes modern-human-like movement of the thumb and index finger in Neanderthals.

The demise of the Neanderthals cannot be attributed to any physical inability to use or manufacture Upper-Palaeolithic-like (Châtelperronian) tools, as the anatomical evidence presented here and the archaeological evidence<sup>10</sup> both indicate that they were capable of manufacturing and handling such implements.

**Wesley A. Niewoehner\***, **Aaron Bergstrom†**, **Derrick Eichele†**, **Melissa Zuroff†**, **Jeffrey T. Clark†**

\*Department of Anthropology, California State University, San Bernardino, California 92407, USA  
e-mail: wniewoeh@csusb.edu

†Archaeology Technologies Lab, North Dakota State University, Fargo, North Dakota 58105, USA

1. Musgrave, J. H. *Nature* **233**, 538–541 (1971).
2. Trinkaus, E. *The Shanidar Neanderthals* (Academic, New York, 1983).
3. Niewoehner, W. A., Weaver, A. & Trinkaus, E. *Am. J. Phys. Anthropol.* **103**, 219–233 (1997).
4. Niewoehner, W. A. *The Functional Anatomy of Late Pleistocene and Recent Human Carpometacarpal and Metacarpophalangeal Articulations* (Univ. New Mexico Press, Albuquerque, 2000).
5. Cooney, W., Lucca, M. J., Chao, E. Y. S. & Linscheid, R. J. *Bone Joint Surg. Am.* **63** A, 1371–1381 (1981).
6. Boule, M. *Ann. Paleontol.* **6**, 111–172; **7**, 21–56, 85–192; **8**, 1–70 (1911–1913); republished by Masson, France (1913).
7. Vleek, E. *Bull. Mém. Soc. Anthropol. Paris* **13**, 257–276 (1975).
8. Kraemer, B. A. & Gilula, L. A. in *The Traumatized Hand and Wrist* (ed. Gilula, L. A.) 65–92 (Saunders, Philadelphia, 1992).
9. Craig, S. M. *Hand Clin.* **8**, 693–700 (1992).
10. Churchill, S. E. & Smith, F. *Yearbook Phys. Anthropol.* **43**, 61–115 (2000).

Competing financial interests: declared none.

## Wetting properties

### When larger drops evaporate faster

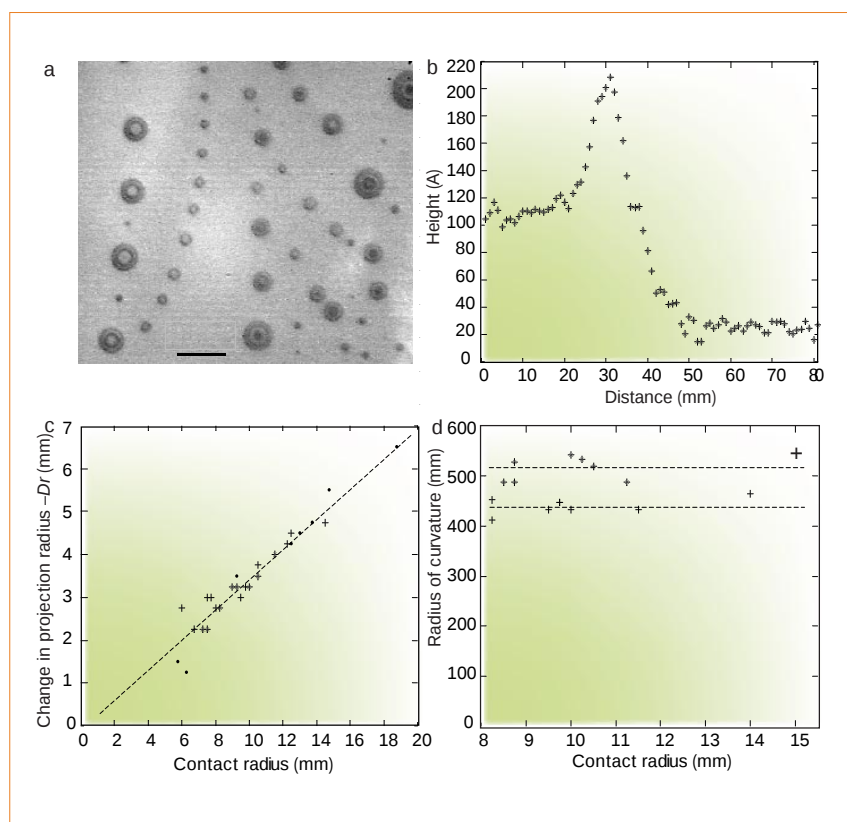
In an ensemble of volatile liquid drops, either in suspension or on a partially wetted substrate, large drops grow at the expense of smaller ones when the pressure is just above the saturated-vapour pressure. Below this critical pressure, all drops evaporate, with the smallest drops evaporating fastest. But an array of water drops of different sizes on a mica substrate cleaved in a vacuum behave differently: below the equilibrium vapour pressure, the largest drops evaporate fastest and the smallest ones more slowly. Here we show that this behaviour is related to the coexistence of thick

and thin water films during evaporation.

An evaporating water film on mica breaks into interesting patterns that have been predicted theoretically<sup>1,2</sup> and have been shown experimentally to result from competing van der Waals and polar surface forces between the water and the substrate<sup>3,4</sup>, with spreading pressures  $S^{\text{vw}}$  and  $S^{\text{p}}$ , respectively. The excess Gibbs free energy per unit area has been described as a function of thickness<sup>3,5,6</sup>, as  $g(h) = S^{\text{vw}}d_0^2/h^2 + S^{\text{p}}\exp[(d_0 - h)/l]$ , where  $l$  is a screening length and  $d_0$  is the molecular diameter.

The resulting interaction potential can have minima at two different film thicknesses, corresponding to molecularly thin and macroscopically thick films. Phase equilibrium between the two films and the





**Figure 1** Behaviour of evaporating water drops on a cleaved mica substrate. **a**, An array of water drops of widely differing sizes at 0 °C, photographed at  $\lambda = 546$  nm using interference contrast. The first dark ring appears at a thickness of 111 nm. Scale bar, 100  $\mu$ m. **b**, Profile of the interface between the thin and thick films, measured by three-beam interferometry, involving reflections at the vapour–film interface and both mica surfaces. **c**, Change in projection radius of drops during 15 min as a function of their radius, for three different experiments under the same conditions,  $p/p_{\text{sat}} = 0.920 \pm 0.002$ . For comparison, on a partially wetted substrate, under the same  $p/p_{\text{sat}}$ , small drops were swallowed by big ones within 30 s. **d**, The radius of curvature of drops is independent of their projection radius. The two lines refer to the same field of drops at different times. Typical error bar is shown on one point.

vapour, corresponding to a Maxwell construction on the chemical potential  $\mu_{\text{film}} = \rho_{\text{liq}}^{-1} dg/dh$ , is similar to that between a liquid and a solid, for example, and a first-order phase transition occurs between them.

The remarkable dendrite-like patterns that develop during evaporation have been shown<sup>4</sup> to mimic two-dimensional diffusion-limited solidification<sup>7,8</sup>. The late-stage pattern consists of an array of slowly evaporating water drops of widely differing sizes (Fig. 1a). Then, if the contact angle is constant, coarsening<sup>6,9</sup> occurs, as in solidification or cooperative evaporation<sup>10</sup>, in which larger drops acquire fluid from smaller ones through vapour transport. A drop with a smaller radius of contact with the substrate has a smaller radius of curvature,  $R$ , and must therefore evaporate faster owing to the Gibbs–Thomson contribution,  $2\gamma/R$ , to its vapour pressure, where  $\gamma$  is the surface tension.

Instead, we have observed that all of the drops continue to evaporate together, the smaller ones even evaporating more slowly than the larger ones. We attribute this behaviour to the thin, continuous film of

water that connects the drops. Figure 1b shows experimental measurements of the thickness of the two films predicted by this model. The two films are separated by a hydrodynamic rim, which has previously been seen both in simulations<sup>3</sup> and in experiments<sup>4,11</sup>. The measurements used a new form of three-beam interferometry (see legend to Fig. 1b).

For the case investigated, the two films have thicknesses of  $25 \pm 5$  and  $110 \pm 10$  Å. The drops, with their free surfaces far from the substrate, have excess chemical potential  $2\rho_{\text{liq}}^{-1}\gamma/R$ . For equilibrium with the continuous thin film,  $R$  must therefore be the same for all the drops. This implies that the contact angle increases with a contact radius. But because the film is continuous, the meeting between liquid, vapour and substrate, which defines the edge of a drop, is not present, and so a conventional contact angle cannot be defined.

It was found in simulations<sup>12</sup> that the contact angle of a drop increases with its volume. Water evaporating from the spherical cap is compensated for by contraction of the contact radius,  $r$ , at velocity  $v = -dr/dt$ . A balance of volumes gives

$2\pi r v (h_2 - h_1) = 2\alpha \rho_{\text{liq}}^{-1} \pi r^2 \gamma / R$ , where  $\alpha$  is the evaporation parameter<sup>4</sup>. As  $R$  is the same for all drops, it can be shown that  $dr/dt$  is proportional to  $-r$ , as has been confirmed experimentally (Fig. 1c, d). The typical situation, in which large drops in an array grow at the expense of the smaller ones, is therefore not universally observed, and when the drops coexist with a microscopically thin continuous surface layer, they behave differently.

**I. Leizerson\*, S. G. Lipson\*, A. V. Lyushnint†**

\*Department of Physics, Technion-Israel Institute of Technology, 32000 Haifa, Israel

e-mail: philya@technion.ac.il

†Department of Theoretical Physics, Perm State Pedagogical University, Perm 614600, Russia

1. Cazabat, A.-M. *Contemp. Phys.* **28**, 347–364 (1987).
2. Brochard-Wyart, F. *Soft Matter Physics* 7–44 (Springer, Berlin, 1995).
3. Sharma, A. & Jameel, A. *J. Colloid Interface Sci.* **161**, 190–208 (1993).
4. Samid-Merzel, N., Lipson, S. G. & Tanhauser, D. S. *Phys. Rev. E* **57**, 2906–2913 (1998).
5. de Gennes, P. G. *Rev. Mod. Phys.* **57**, 325–359 (1985).
6. Israelachvili, J. *Intermolecular and Surface Forces* 2nd edn (Academic, San Diego, 1992).
7. Ben-Jacob, E. *Nature* **343**, 523–530 (1990).
8. Ihle, T. & Muller-Krumbhaar, H. *Phys. Rev. Lett.* **70**, 3083–3086 (1993).
9. McHale, G., Rowan, S. M., Newton, M. I. & Banerjee, M. K. *J. Phys. Chem. B* **102**, 1964–1967 (1998).
10. Schaffle, C., Bechinger, C., Rinn, B., David, C. & Leider, P. *Phys. Rev. Lett.* **83**, 5302–5305 (1999).
11. Elbaum, M. & Lipson, S. G. *Phys. Rev. Lett.* **72**, 3562–3565 (1994).
12. Sharma, A. *Langmuir* **9**, 3580–3586 (1993).

**Competing financial interests:** declared none.

## Palaeobotany

# Swimming sperm in an extinct Gondwanan plant

**T**he now-extinct plant *Glossopteris* that dominated the Southern Hemisphere (Gondwana) during the Permian period serves as early evidence of continental drift<sup>1,2</sup>, and may be ancestral to the group of seed plants known as angiosperms<sup>3</sup>. Here we describe a 250-million-year-old fossil from Homevale in Queensland, Australia, of anatomically preserved pollen tubes and swimming male gametes from *Glossopteris*. The discovery of this simple reproductive system in *Glossopteris* has implications for its phylogenetic relationships with extant groups of seed plants (conifers and flowering plants, for example) and for the evolution of pollination biology in general.

Five fossilized pollen tubes are evident in the pollen chamber of a single ovule, which has been attributed to *Glossopteris homevalensis*<sup>2</sup> (Fig. 1a, b). They are short and ovoid, unbranched, and contain gametes in several stages of development. Two of the pollen tubes contain a single, immature, spermatogenous cell, two of

# Is fertilization efficiency misleading?

Given the need to increase crop production in the future while minimizing any associated impact on the environment, it is important to understand the relationship between global crop production and fertilization. Tilman *et al.*<sup>1</sup> argue that fertilization beyond current levels is unlikely to increase crop yields (production per unit area) as effectively as in the past, due to diminishing returns. However, their evidence is misleading and does not support their conclusions. Diminishing returns are not readily apparent on a global scale.

Figure 2 of Tilman *et al.*<sup>1</sup> shows trends in global cereal yield and 'nitrogen-fertilization efficiency' (defined as production per unit mass of nitrogen-fertilizer consumption) from 1961 to 1996. In the figure, nitrogen-fertilization efficiency plummets over time while yield climbs relatively linearly, presumably as fertilization increased over this period (verified here using data from the United Nations Food and Agriculture Organization). The figure seemingly supports the authors' statements, such as "further increases in nitrogen and phosphorus application are unlikely to be as effective at increasing yields (Fig. 2a) because of diminishing returns (Fig. 2b)". However, the response depicted in their Fig. 2b is a mathematical artefact of the relationship between crop yield and fertilizer-application rate (mass per unit area), and so does not support the authors' conclusions.

In fertilizing crops, the ideal response would be one in which a given increase in fertilizer-application rate leads to a constant increase in crop yield or, identically, a given increase in total fertilizer application leads to a constant increase in production. This ideal system, in which the effectiveness of fertilization does not decline as fertilizer use increases, is mathematically represented by  $p = af + b$  (where  $p$  is the crop production,  $f$  is the amount of fertilizer applied,  $a$  is the slope of the response and  $b$  is the production in the absence of fertilizer).

With this linear response, the nitrogen-fertilization efficiency is represented by  $p/f = a + b/f$ , and so is inversely proportional to  $f$ . As  $f$  approaches zero, this efficiency approaches infinity, and declines as  $f$  increases, tending towards the value of  $a$ . This behaviour is quantified by the derivative of fertilization efficiency, namely  $-b/f^2$ . Thus, even when a crop shows an ideal response to fertilization, fertilization efficiency declines (Fig. 1).

It seems that global cereal production has followed the linear relationship described above, and has not shown dimin-

multilayered structure reaches 300  $\mu\text{m}$  in length, with 1,000 flagella on three gyres<sup>4</sup>, whereas that of the cycad *Zamia*, one of the largest known plant sperm, is a helical band of 1,600  $\mu\text{m}$  in length, with six gyres bearing several tens of thousands of flagella arranged in 8–12 rows at short intervals (0.1  $\mu\text{m}$ ; ref. 5).

*Glossopteris* sperm is therefore more similar in size to that of most other vascular plants, including pteridophytes (*Botrychium*, 10  $\mu\text{m}$ , ref. 6; *Lycopodium*, 8–10  $\mu\text{m}$ , ref. 7; *Equisetum*, 19.7  $\mu\text{m}$ , ref. 7), conifers (*Torreya*, 13–19.5  $\mu\text{m}$ ; *Pinus*, 24  $\mu\text{m}$ ; *Thuja*, 35  $\mu\text{m}$ ; ref. 10), gnetophytes (*Ephedra*, 16  $\mu\text{m}$ ; ref. 10) and angiosperms (*Hordeum*, 11.7  $\mu\text{m}$ ; ref. 11). Only the putative sperm found in fossil medullosans are comparable in size to those of cycads, to which they are often allied, and whose sperm are unusually large<sup>9,12</sup>.

The pollen tube of *Glossopteris* differs from the highly branched, haustorial pollen tubes of both *Ginkgo* and cycads, as well as from the longer, sparsely branched or unbranched siphonogamous pollen tubes of *Callistophyton*, conifers, bennettitaleans<sup>13</sup> and angiosperms. *Glossopteris* pollen tubes are the simplest known among vascular plants. They are slightly elongated, unlike in the fossil medullosan ovule *Pachytesta*, where mature sperm inside the pollen protoplast apparently did not elongate to form a pollen tube<sup>12</sup> — a type of structure that has been interpreted as an early stage in the evolution of pollen tubes<sup>12</sup>. Together, our findings indicate that *Glossopteris* had a very simple mode of reproduction that was more similar to that of cycads and *Ginkgo* than to those of other extant seed plants.

Harufumi Nishida\*, Kathleen B. Pigg†, John F. Rigby‡

\*Faculty of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo, Tokyo 112-8551, Japan

e-mail: nishida@mechgate.mech.chuo-u.ac.jp

†Department of Plant Biology, Box 871601, Arizona State University, Tempe, Arizona 85287-1601, USA

‡School of Natural Resource Sciences, Queensland University of Technology, Box 2434, Brisbane, Queensland 4001, Australia

1. Gould, R. E. & Delevoryas, T. *Alcheringa* **1**, 387–399 (1977).

2. Pigg, K. B. & McLoughlin, S. *Rev. Palaeobot. Palynol.* **97**, 339–359 (1997).

3. Retallack, G. & Dilcher, D. L. *Paleobiology* **7**, 54–67 (1988).

4. Li, Y. *et al.* *Protoplasma* **149**, 57–63 (1989).

5. Norstog, K. J. *Bot. Gaz.* **147**, 40–46 (1986).

6. Bierhorst, D. W. *Morphology of Vascular Plants* (Macmillan, New York, 1971).

7. Gifford, E. M. & Foster, A. S. *Morphology and Evolution of Vascular Plants* 3rd edn (Freeman, New York, 1989).

8. Hirase, S. *Bot. Mag. Tokyo* **10**, 325–328 (1896).

9. Ikeno, S. *Bot. Mag. Tokyo* **10**, 367–368 (1896).

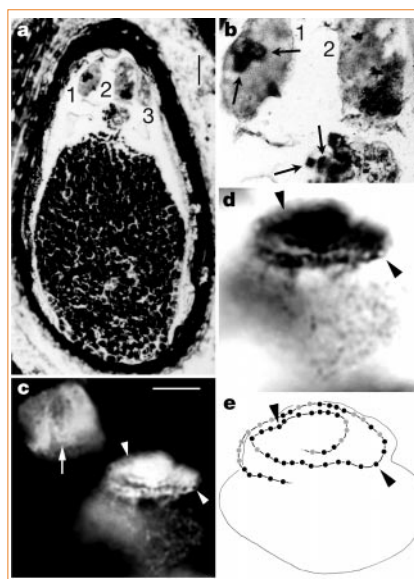
10. Chamberlain, C. J. *Gymnosperms* (Univ. Chicago Press, 1935).

11. Cass, D. & Jensen, W. A. *Am. J. Bot.* **57**, 2–70 (1970).

12. Stewart, W. N. *Am. Midl. Natur.* **46**, 717–742 (1951).

13. Rothwell, G. W. & Stockey, R. A. *Botany 2001 Abstracts* **25** (2001).

Competing financial interests: declared none.

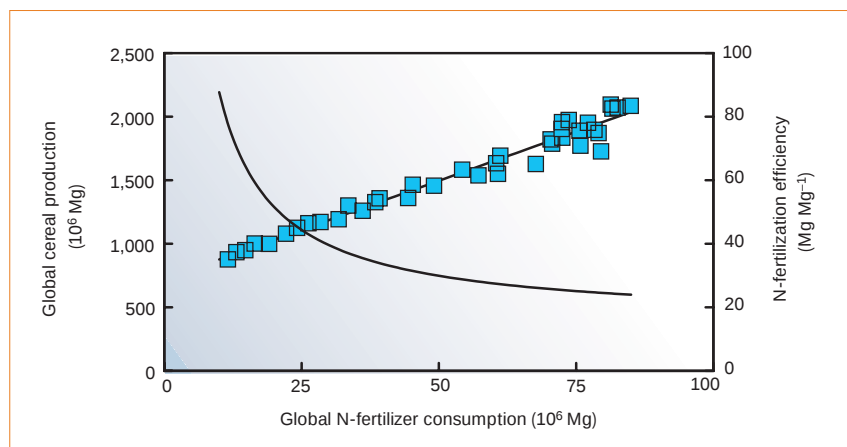


**Figure 1** Fossil motile sperm in a Late Permian *Glossopteris* ovule. Specimens are housed at Chuo University, Tokyo. **a–d**, Light micrographs of specimen H397018 E<sub>2</sub>lat., slide 5. **a**, Longitudinal section of the ovule, showing the megagametophyte (below) and three pollen tubes (numbered 1–3) in the pollen chamber (top). **b**, Enlarged view of pollen tubes 1 and 2. Top-left arrows indicate immature sperm in tube 1. Tube 2 (right) is ruptured basally and two motile sperm have been released (lower-right arrows; also evident are the contents of tube 2 and megagametophyte tissue). **c**, The two released sperm at higher magnification; image has been reversed to enhance detail. Arrow, multilayered helical structure represented by alignment of circular bodies; arrowheads, line of basal bodies. **d**, Detail of sperm on the right in **c**, showing the basal-body alignment (arrowheads). **e**, Schematic diagram of the same sperm, showing the positioning of the basal bodies (arrowheads). Scale bars, 100  $\mu\text{m}$  (**a**, **b**) and 5  $\mu\text{m}$  (**c**, **d**).

them each contain a pair of young sperm of immature morphology, and the fifth has been ruptured basally, evidently in the process of dispersing two mature sperm (Fig. 1b–d). No prothallial or sterile cells are evident.

One entire sperm is about 13.9  $\mu\text{m}$  across, with a helical configuration that represents the multilayered structure upon which numerous basal bodies are inserted<sup>4,5</sup>. This structure is strikingly similar to the flagellate band of many living pteridophytes<sup>6,7</sup>, *Ginkgo*<sup>8</sup> and cycads<sup>9</sup>. Part of a second sperm also has a spiral structure (Fig. 1c, left arrow). A single line of small dots occurs at roughly regular intervals (Fig. 1c, d, arrowheads), which are inferred to be basal bodies or the basal plate that connects the basal bodies to the flagella (Fig. 1e).

A mature *Glossopteris* sperm has 37 basal bodies, as determined by optical sectioning. A complete multilayered structure is estimated to be about 45  $\mu\text{m}$  long with more than 50 basal bodies, each at a mean interval of 0.57  $\mu\text{m}$ , which form two anticlockwise gyres (Fig. 1e). By contrast, *Ginkgo*'s



**Figure 1** Regression between global cereal production and nitrogen-fertilizer consumption, showing a relatively constant increase in production per unit increase in fertilizer use, and no evidence of diminishing returns. Boxed points are actual data (from the United Nations Food and Agriculture Organization 2002; <http://apps.fao.org>); the thin line ( $P = 15.47f + 722.3$ ) is the result of a regression of production on fertilizer consumption. Even within this linear relationship, nitrogen-fertilization efficiency (calculated by dividing the regression line by fertilizer consumption; thick line) declines as fertilizer consumption increases.

ishing returns from fertilization (Fig. 1; data from the United Nations Food and Agriculture Organization). The term 'diminishing returns', which was first used in relation to agricultural productivity<sup>2,3</sup>, describes a decline in the marginal output, or the increase in output per unit change in input, when the level of a variable input is increased<sup>4</sup>, rather than a decline in the ratio of output to input. 'Nitrogen-fertilization efficiency', as calculated by Tilman *et al.*<sup>1</sup>, is not a useful way to represent the effectiveness of fertilization.

**Sasha D. Hafner**

Department of Environmental and Forest Biology,  
State University of New York, College of  
Environmental Science and Forestry, Syracuse,  
New York 13210, USA  
e-mail: [sdhafner@syr.edu](mailto:sdhafner@syr.edu)

1. Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R. & Polasky, S. *Nature* **418**, 671–677 (2002).
2. Sraffa, P. & Dobb, M. H. (eds) *The Works and Correspondence of David Ricardo* Vol. 3, 1809–1811 (Cambridge Univ. Press, Cambridge, 1951).
3. Sraffa, P. & Dobb, M. H. (eds) *The Works and Correspondence of David Ricardo* Vol. 4, 1815–1823 (Cambridge Univ. Press, Cambridge, 1951).
4. Case, K. E. & Fair, R. C. *Principles of Microeconomics* 4th edn (Prentice Hall, New Jersey, 1996).

**Tilman *et al.* reply** — An important question that arises from Hafner's comment is why global crop production<sup>1</sup> should appear to be a linear function of global nitrogen fertilization, even though crop production has consistently shown diminishing returns from increased fertilization in field trials<sup>2–4</sup>. Although we concur that our Fig. 2b does not unambiguously demonstrate diminishing returns, evidence based on data from the United Nations Food and Agriculture Organization indicate that global crop production and yield<sup>5</sup> do indeed show diminishing returns for increased fertilization.

Crop responses to nitrogen fertilizer are expected to exhibit diminishing returns because yields may be limited simultaneously by the availability of light, water, nitrogen and the other essential nutrients. As nitrogen fertilizer is added to alleviate nitrogen deficiency, other resources will then become limiting, causing the crop's response to nitrogen to diminish<sup>6</sup>. Moreover, nitrogen losses increase with increased nitrogen application because surplus inorganic nitrogen ( $\text{NO}_3^-$  and  $\text{NH}_4^+$ ) in the root zone is leached into the groundwater or lost to the atmosphere in gaseous form.

Hafner's univariate analysis ignores other agricultural inputs, particularly irrigation, that change simultaneously with increasing nitrogen fertilization. It is more appropriate to determine the dependence of global crop production (cereals, coarse grains and root crops) and global crop yield (production divided by global land dedicated to these crops) on global nitrogen fertilization while controlling for global irrigation.

In tests for a nonlinear (saturating) effect of nitrogen fertilization (J. Fargione *et al.*, unpublished results), one using global rates of production, fertilization and irrigation per hectare, and the other using production, fertilization and irrigation per hectare, we found a significantly saturating effect of nitrogen fertilization. Multiple regression of yearly global crop production on irrigation and nitrogen fertilization revealed a significant positive linear effect of irrigation ( $F_{1,36} = 56.9$ ,  $P < 0.0001$ ) and a simultaneous saturating (quadratic) effect of nitrogen fertilization. The nitrogen effect had a significantly positive linear ( $F_{1,36} = 13.8$ ,

$P = 0.0007$ ) and a significantly negative quadratic ( $F_{1,36} = 7.23$ ,  $P < 0.018$ ) term for 1961–2000. The negative quadratic term in the fitted model gave a curve for the dependence of global crop production on nitrogen fertilization that approached its peak at current global rates of nitrogen fertilization.

A multiple regression of global crop yield (production per hectare) on global irrigation per hectare and on global nitrogen fertilization per hectare also indicated diminishing returns (J. Fargione *et al.*, unpublished results). Global crop yield increased with irrigation per hectare ( $F_{1,36} = 64.2$ ,  $P < 0.0001$ ) but was a saturating function of fertilization per hectare, with a positive linear term ( $F_{1,36} = 13.4$ ,  $P = 0.0008$ ) and a negative quadratic term ( $F_{1,36} = 4.36$ ,  $P = 0.044$ ) for this 40-year period. The fitted curve for yield approached its peak at current global rates of nitrogen fertilization. Twenty-two similar analyses for each of 11 periods from 1961 through to each year from 1990 to 2000 also revealed significantly diminishing returns.

Although multiple regressions that use collinear variables cannot demonstrate causal relationships unambiguously, and although we did not control for improved crop varieties, it is likely that there have been diminishing returns of increased global nitrogen fertilization at least since 1990. If this is the case, other technologies (see ref. 7, for example) may help to increase global crop yields; indeed, the United States' maize yield increased by almost 40% from 1980–2000 without any increase in nitrogen fertilization<sup>6</sup>. Plant breeding, biotechnology and advances in crop and soil management will probably account for most of the future increases in global crop production, without the negative environmental effects that are attributed to nitrogen fertilizers.

**David Tilman\*, Kenneth G. Cassman, Pamela A. Matson, Rosamond Naylor, Stephen Polasky**

\*Department of Ecology, Evolution and Behavior,  
University of Minnesota, St Paul,  
Minnesota 55108, USA  
e-mail: [tilman@umn.edu](mailto:tilman@umn.edu)

1. Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R. & Polasky, S. *Nature* **418**, 671–677 (2002).
2. Willcox, O. W. J. *Am. Soc. Agron.* **36**, 20–31 (1944).
3. Dahnke, W. C. & Olson, R. A. in *Soil Testing and Plant Analysis* (ed. Westerman, R. L.) 45–71 (Soil Sci. Soc. Am., New York, 1990).
4. Hayami, Y. & Ruttan, V. W. *Agricultural Development: An International Perspective* (Johns Hopkins Univ. Press, Baltimore, 1985).
5. United Nations Food and Agriculture Organization Statistical Databases, <http://apps.fao.org> (2002).
6. Cassman, K. G., Dobermann, A. & Walters, D. *Ambio* **31**, 132–140 (2002).
7. Ruttan, V. W. *Proc. Natl Acad. Sci. USA* **96**, 5960–5967 (1999).

**brief communications** is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on *Nature's* website ([www.nature.com/nature](http://www.nature.com/nature)) or on request from [nature@nature.com](mailto:nature@nature.com).



# Three-dimensional structural dynamics of myosin V by single-molecule fluorescence polarization

Joseph N. Forkey\*, Margot E. Quinlan\*, M. Alexander Shaw\*, John E. T. Corrie† & Yale E. Goldman\*

\* Pennsylvania Muscle Institute, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6083, USA

† National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

**The structural change that generates force and motion in actomyosin motility has been proposed to be tilting of the myosin light chain domain, which serves as a lever arm. Several experimental approaches have provided support for the lever arm hypothesis; however, the extent and timing of tilting motions are not well defined in the motor protein complex of functioning actomyosin. Here we report three-dimensional measurements of the structural dynamics of the light chain domain of brain myosin V using a single-molecule fluorescence polarization technique that determines the orientation of individual protein domains with 20–40-ms time resolution. Single fluorescent calmodulin light chains tilted back and forth between two well-defined angles as the myosin molecule processively translocated along actin. The results provide evidence for lever arm rotation of the calmodulin-binding domain in myosin V, and support a ‘hand-over-hand’ mechanism for the translocation of double-headed myosin V molecules along actin filaments. The technique is applicable to the study of real-time structural changes in other biological systems.**

Force and translocation in actomyosin systems are thought to be generated by motions in which the myosin light chain domain (LCD) acts as a mechanical lever arm. In the lever arm hypothesis, the amino-terminal motor domain of myosin binds stereospecifically to actin while the LCD rotates relative to the motor domain<sup>1,2</sup>. The motor domain and LCD can occupy a number of relative orientations<sup>3,4</sup>, and LCD rotations accompany translocation and force generation in muscle fibres<sup>5,6</sup>. Despite extensive work, however, time-resolved angle changes of the LCD with appropriate magnitude to explain observed stepping distances and velocities have not yet been demonstrated clearly in an active system.

X-ray crystallography and cryo-electron microscopy provide high spatial resolution but they cannot measure changes in a functioning sample over time. On the other hand, techniques such as spectroscopy and low-angle X-ray diffraction of bulk samples, such as muscle fibres, have provided good temporal resolution, but are difficult to quantify and interpret structurally because averaging among unsynchronized populations of molecules reduces the signal changes. To circumvent these difficulties, we have developed a single-molecule fluorescence polarization technique, which is able to determine the orientation of individual protein domains to a time resolution of 20–40 ms. This method detects the mean three-dimensional orientation as well as the amplitude of wobble occupying the timescale  $\gg 4$  ns to  $\ll 10$  ms. Here we report its use to observe, in real time, large angle changes in the LCD of individual myosin V molecules, with magnitude and dynamics consistent with the lever arm hypothesis.

We used myosin V for this study because its specific enzymatic and mechanical features<sup>7–10</sup> confer advantages for single-molecule fluorescence polarization measurements. Individual double-headed myosin V molecules move along actin with high processivity<sup>11,12</sup>, undergoing many biochemical cycles and multiple 36–37-nm mechanical steps for each encounter with an actin filament. A myosin V molecule frequently moves along an actin filament for several seconds, allowing the determination of its velocity followed by measurement of orientation changes during several catalytic cycles. The most widely held hypothesis that explains myosin V processivity is the so-called ‘hand-over-hand’ model, which predicts that each head spends most of its ATPase cycle attached to the actin

filament, either leading or trailing its partner head. Stepping events are thought to occur through detachment of the trailing head, rapid forward motion of this head by about 74 nm, and reattachment as the new leading head, while the other head remains attached<sup>8,11,13–15</sup>. The translation of the detached head from trailing to leading position is thought to be generated mainly by tilting of the LCD of the attached head. On the basis of this model, we expected to observe two discrete orientations of the myosin V LCD with similar lifetimes, corresponding to the two head positions.

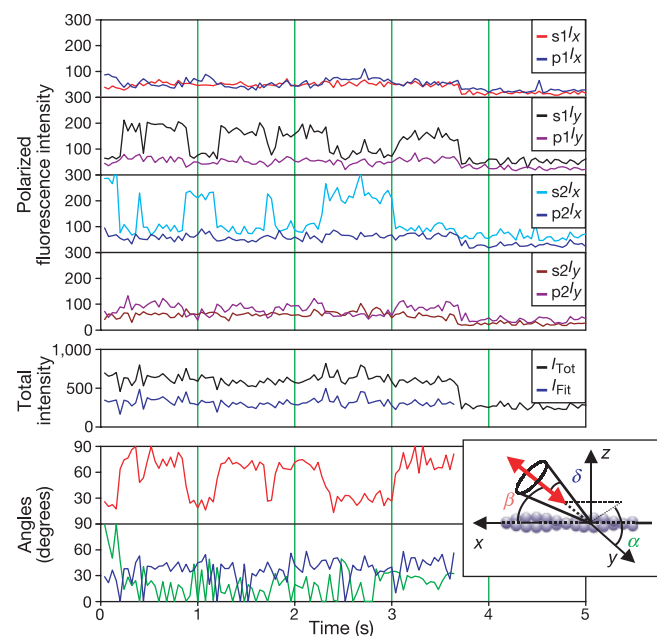
## Time-resolved single-molecule fluorescence polarization

Calmodulin (CaM) was labelled with bisiodoacetamidorhodamine<sup>16,17</sup> at a known orientation (see Supplementary Information) relative to the molecular structure and exchanged for endogenous CaM on myosin V at low stoichiometry (on average  $<0.5$  labelled CaM light chains per double-headed myosin V molecule). Labelled myosin V molecules in a sample chamber on a fused silica microscope slide diffused from the bulk medium to interact with biotin- and AEDANS-labelled F-actin, which was immobilized on the slide surface (see Methods). Total internal reflection fluorescence microscopy<sup>18,19</sup> was used to image alternately the AEDANS-actin and rhodamine-labelled myosin V at the slide surface with evanescent waves at two distinct wavelengths. Individual rhodamine-labelled myosin V molecules, which were localized with an actin filament and moved along it when ATP was present, were selected for further investigation. These molecules were excited by an evanescent wave with four different polarization states switched (time-multiplexed) at intervals of 5 or 10 ms each, in aggregate filling out the three orthogonal spatial directions of optical polarization ( $x$  and  $y$  in the plane of the microscope slide, and  $z$  along the microscope optical axis; see Methods). The fluorescence was resolved into its components linearly polarized along the  $x$  and  $y$  axes, and was detected by two photon-counting avalanche photodiodes. The four time-multiplexed excitation and two simultaneously detected emission polarizations resulted in eight polarized fluorescence intensities for each successive 20- or 40-ms time point.

Fluorescence traces from rhodamine-labelled myosin V molecules containing single fluorophores were identified by a characteristic

intensity as well as simultaneous discrete photobleaching of all polarized intensities to background levels (Figs 1 and 2). Data were obtained from myosin V molecules interacting with actin filaments at 0 and 5  $\mu\text{M}$  ATP with 40-ms time resolution, and at 40  $\mu\text{M}$  ATP with 20-ms time resolution. Many of the molecules at 5  $\mu\text{M}$  ATP (Fig. 1) and at 40  $\mu\text{M}$  ATP (Supplementary Information) exhibited polarized fluorescence intensity levels that repeatedly underwent transitions back and forth in opposite directions. In Fig. 1, these transitions are most apparent in the intensities labelled  $s_1I_y$  and  $s_2I_x$  (see Methods for definitions). A weighted sum of all eight intensities ( $I_{\text{Tot}}$ ), representing the total fluorescence independent of orientation, was constant, indicating that the intensity changes of the individual traces were due to abrupt rotations of the fluorophore. As the dual covalent linkage between the fluorophore and CaM fixes their relative orientation, these results imply that the LCD of myosin V tilts back and forth. For each time point in the single-molecule trace, the polarized intensities were fit to a set of equations to determine the total intensity ( $I_{\text{Fit}}$ ), the axial angle ( $\beta$ ) of the fluorophore relative to the actin axis, the azimuthal angle ( $\alpha$ ) around that axis, and the half-angle ( $\delta$ ) of a cone describing the wobbling motion of the fluorophore<sup>20</sup> on a timescale much slower than 4 ns but much faster than 10 ms (Fig. 1, lower panels and inset). The abrupt jumps of fluorescence intensity are caused by large changes in the  $\beta$ -value between periods of relatively steady  $\beta$ , as predicted by the hand-over-hand model.

To clarify the time course of these tilting motions relative to the noise in the recordings, individual transitions from many molecules were effectively synchronized by shifting their time bases to align the start times, and then averaged. The average angular traces derived from this procedure (Fig. 3; see also Supplementary Information) show that  $\beta$ -values changed by 40–50° and that this change was



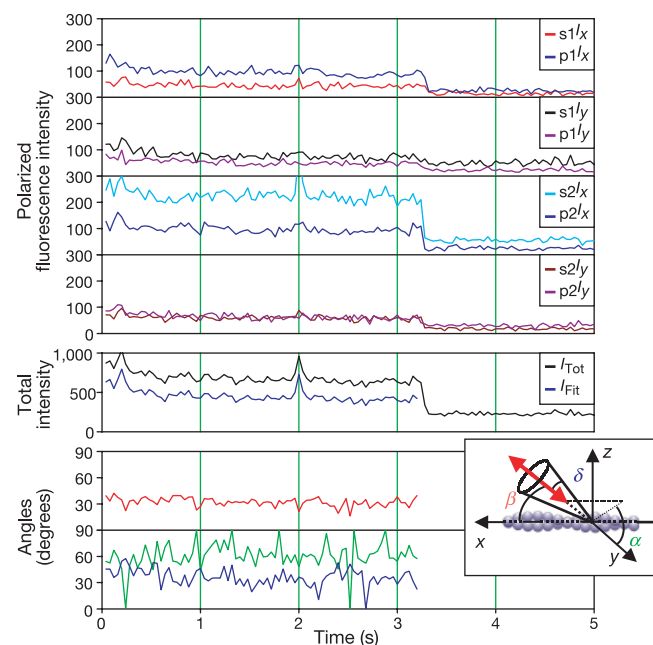
**Figure 1** Data trace from a single myosin V molecule translocating along actin in the presence of 5  $\mu\text{M}$  ATP. Polarized fluorescence intensities have units of photocounts per 10-ms gate; total intensities have units of photocounts per 40-ms cycle. Polarized intensity subscripts indicate excitation/detection polarizations as defined in the Methods.  $I_{\text{Fit}}$  is the total intensity determined by fitting the polarized fluorescence intensities;  $I_{\text{Tot}}$  is calculated from the polarized intensities as described in the Methods and should approximate  $I_{\text{Fit}}$  plus background. The single discrete decrease of all intensities to background levels at 3.7 s is due to photobleaching of the fluorophore. The inset defines  $\beta$ ,  $\alpha$  and  $\delta$  (red, green and blue traces, respectively, in bottom panels) relative to the actin filament and the microscope frame, which is defined with the stage in the  $x$ - $y$  plane.

abrupt relative to the 20–40-ms recording time resolution. Furthermore, the averaging procedure revealed a step change of 15–20° in  $\alpha$ . Whether this angle change implies left- or right-handed pitch in the step is not discernable from the current data owing to symmetries of the probe dipole and the experimental setup. The mobility parameter,  $\delta$ , averaged 30–35° and changed much less than  $\alpha$  or  $\beta$ . A small spike in  $\delta$  at the transition is probably a sampling artefact due to limited temporal resolution and time-multiplexing of the excitation polarizations. These results imply that, at the ATP concentrations studied, any intermediate state with a different LCD orientation, such as that associated with a detached head, can occupy at most 10% of the average cycle time. This result is in agreement with a hand-over-hand model in which both heads of myosin V are attached to actin for most of the ATPase cycle.

Traces from a subset of molecules observed with 5 or 40  $\mu\text{M}$  ATP, and nearly all traces recorded at 0 ATP (with apyrase present, Fig. 2), showed constant intensities, corresponding to steady angles. Angular transitions were observed at 0 ATP only rarely (0.1 events per molecule as compared with 3.1 and 2.4 for 5 and 40  $\mu\text{M}$  ATP, respectively), consistent with occasional photochemical processes, the details of which will be reported elsewhere (M.E.Q., J.N.F. and Y.E.G., manuscript in preparation). None of the traces at 0 ATP showed the reversible and repeated transitions observed at 5 and 40  $\mu\text{M}$  ATP, indicating that the orientation of a myosin V molecule bound to actin was stable in the absence of ATP.

### Angular distributions

To characterize further the LCD orientations, an average  $\beta$ -value was calculated over each interval of the traces where the  $\beta$ -value was steady. This procedure resulted in many average  $\beta$ -values for each molecule that exhibited tilting at 5 and 40  $\mu\text{M}$  ATP, but only one (or occasionally two) average  $\beta$ -value for each molecule that did not exhibit tilting at 0, 5 and 40  $\mu\text{M}$  ATP. Distributions of average  $\beta$ -values for these groups of molecules showed discrete peaks that were fitted by gaussian-shaped curves (Fig. 4, Table 1 and Supplementary Information). An angular component with peak  $\beta$  at 27–30° and a relatively small wobble amplitude ( $\delta_1 = 32$ –36°) was detected for molecules in the absence of ATP and for molecules exhibiting tilting at 5 and 40  $\mu\text{M}$  ATP. Dispersion of these angular



**Figure 2** Data trace from a myosin V molecule bound to actin in the absence of ATP. All traces are relatively flat except for noise until the probe bleaches at 3.3 s.

components,  $\sigma_{\beta 1}$  (Table 1), was lower at 0 ATP, reflecting the well-defined orientation of an attached myosin head near the end of its working stroke in the absence of ATP. Thus, it is likely that this angular component (at  $\beta_1 = 27\text{--}30^\circ$ ) also represents the end of the working stroke in the active case. In the hand-over-hand mechanism, this corresponds to the trailing head with its LCD tilted forward towards the barbed end of the actin filament.

An angular component with a higher  $\beta$ -value, peaking at  $72\text{--}78^\circ$  for the tilting molecules at 5 and  $40\text{ }\mu\text{M}$  ATP, would then represent the leading heads. Owing to symmetries of the probe dipole and the experimental setup, it is possible that the two  $\beta$ -peaks are actually on opposite sides of the plane perpendicular to actin (that is, one peak is at  $180^\circ - \beta$ ). As discussed later, the peak at  $\tilde{\beta}_2 = 72\text{--}78^\circ$  for the tilting molecules probably represents probes at  $108\text{--}102^\circ$ .

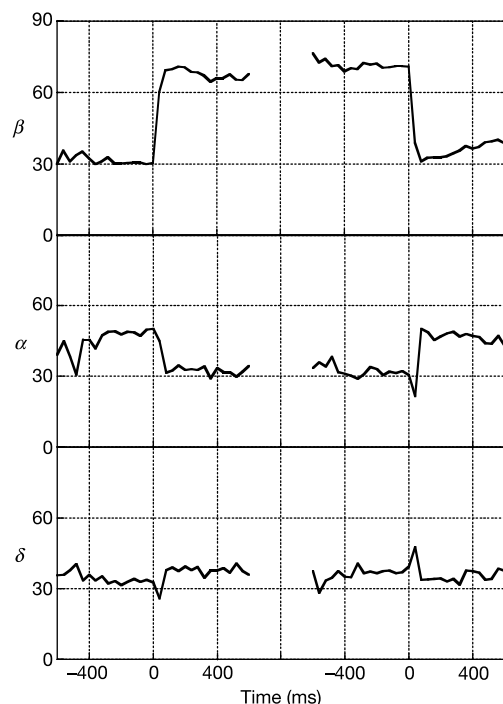
For the set of molecules that did not tilt in the presence of ATP, the average  $\beta$ -distributions peaked at  $52\text{--}54^\circ$  and the mobility ( $\tilde{\delta}_2$ ) was  $42\text{--}45^\circ$  (Fig. 4b, Table 1 and Supplementary Information). These values match a component in the 0 ATP distribution with similar  $\beta$ - and  $\delta$ -values, suggesting that both correspond to the same population of labelled CaM light chains. The absence of tilting may indicate that these CaM light chains are bound to a different site on myosin V than those that display tilting. If a labelled CaM were bound to a myosin V molecule outside the LCD, it would not be expected to tilt. Another potential explanation for non-tilting CaM light chains is that the LCD of the leading head may bend after attaching to actin, so that the section of LCD closest to the motor domain retains the same orientation in both leading and trailing positions. In that case, a CaM bound to this proximal region of the LCD would not exhibit tilting. It is also possible that non-tilting traces represent CaM light chains associated with myosin V heads that have stopped moving. Such heads could be detached from actin, but remain close to the surface if, for instance, their partner heads were still bound to actin. For molecules in the absence of ATP, double-headed attachments are unlikely<sup>13</sup>, so the two components

in Fig. 4c may represent attached ( $\tilde{\beta}_1 = 27^\circ$ ,  $\tilde{\delta}_1 = 36^\circ$ ) and detached ( $\tilde{\beta}_2 = 51^\circ$ ,  $\tilde{\delta}_2 = 46^\circ$ ) heads.

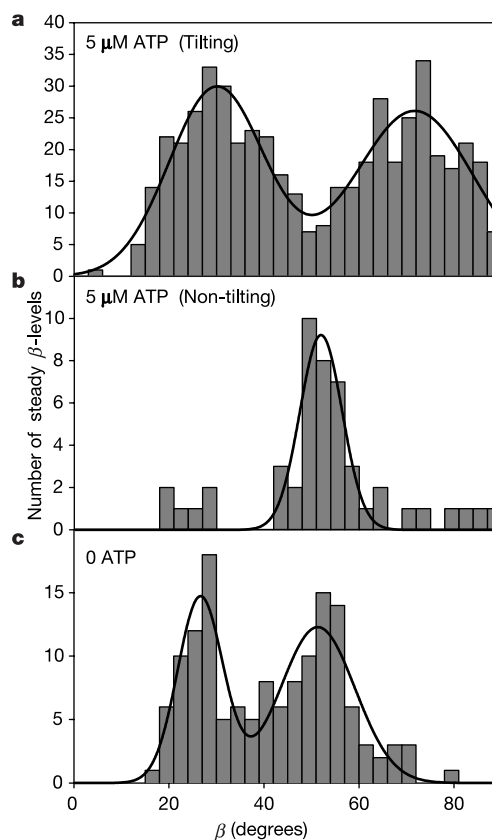
### Kinetics of tilting and translocation

For each molecule that exhibited tilting, structural events containing an interval of steady  $\beta$  flanked by discrete transitions were divided into two groups with average  $\beta$ -values either greater or less than  $50^\circ$ . Event dwell times were measured as the interval between the beginnings of consecutive events. Histograms of these dwell times showed approximately exponential decay with no significant difference in rate between events at high and low  $\beta$ -values (see Methods).

Histograms of aggregate dwell time including both high and low  $\beta$ -events are shown in Fig. 5. The event durations decreased markedly when the ATP concentration was raised from 5 to  $40\text{ }\mu\text{M}$ . The rate constant for a single exponential decay fitted to each of these histograms (dashed lines) increased 2.7-fold for the 8-fold increase in ATP concentration (see Methods), indicating that the dwell times are determined by both an ATP-dependent kinetic process and an ATP-independent process. Therefore, a double exponential curve assuming two sequential processes, one with rate  $k_1[\text{ATP}]$ , and the other with rate  $k_2$ , independent of  $[\text{ATP}]$ , was fitted to the data in both histograms (solid lines in Fig. 5a, b). This global analysis yielded rate constants of  $k_1 = 1.1 \pm 0.2 \times 10^6\text{ M}^{-1}\text{ s}^{-1}$ , and  $k_2 = 12.0 \pm 1.7\text{ s}^{-1}$  (95% confidence limits). These values are in good agreement with the ATP-binding rate of  $0.9 \times 10^6\text{ M}^{-1}\text{ s}^{-1}$  and the ADP release rate of  $12\text{--}16\text{ s}^{-1}$  measured biochemically<sup>8</sup>, and with the two rate constants ( $0.9 \times 10^6\text{ M}^{-1}\text{ s}^{-1}$



**Figure 3** Transition averages. Point-by-point averages of 208 tilting events from 86 molecules in the presence of  $5\text{ }\mu\text{M}$  ATP were determined after synchronizing the transitions at the points where the  $\beta$ -value abruptly increased or decreased. The time course shows that the tilting is rapid and incorporates axial ( $\beta$ ) and azimuthal ( $\alpha$ ) components.



**Figure 4** Distributions of average  $\beta$ -values in the presence and absence of ATP. **a**, Myosin V molecules translocating and showing repeated tilting events at  $5\text{ }\mu\text{M}$  ATP. **b**, Molecules translocating in the presence of  $5\text{ }\mu\text{M}$  ATP but not showing angle changes. **c**, Myosin V co-localized with actin filaments in the absence of ATP. Solid lines indicate the best fits to single or double gaussian-shaped profiles. Parameters of the fits are given in Table 1.



Table 1 Peak angles, dispersions and mobilities

| ATP concentration      | <i>n</i> | Fraction in first peak | $\bar{\beta}_1$ | $\sigma_{\beta 1}$ | $\bar{\delta}_1$ | $\sigma_{\delta 1}$ | $\bar{\beta}_2$ | $\sigma_{\beta 2}$ | $\bar{\delta}_2$ | $\sigma_{\delta 2}$ |
|------------------------|----------|------------------------|-----------------|--------------------|------------------|---------------------|-----------------|--------------------|------------------|---------------------|
| 5 $\mu$ M tilting      | 86       | 0.48 $\pm$ 0.04        | 30.0 $\pm$ 0.9  | 10.0 $\pm$ 0.9     | 34.4             | 13.6                | 71.7 $\pm$ 1.1  | 12.4 $\pm$ 1.5     | 37.3             | 11.2                |
| 40 $\mu$ M tilting     | 100      | 0.46 $\pm$ 0.03        | 27.1 $\pm$ 0.6  | 8.92 $\pm$ 0.6     | 31.6             | 11.4                | 77.8 $\pm$ 1.3  | 13.1 $\pm$ 1.5     | 30.4             | 12.1                |
| 5 $\mu$ M non-tilting  | 48       | —                      | —               | —                  | —                | —                   | 52.0 $\pm$ 0.5  | 4.5 $\pm$ 0.5      | 44.9             | 9.0                 |
| 40 $\mu$ M non-tilting | 101      | —                      | —               | —                  | —                | —                   | 54.2 $\pm$ 0.7  | 6.2 $\pm$ 0.7      | 41.9             | 8.3                 |
| 0 $\mu$ M              | 140      | 0.42 $\pm$ 0.04        | 26.6 $\pm$ 0.6  | 4.78 $\pm$ 0.6     | 36.1             | 7.6                 | 51.3 $\pm$ 0.9  | 7.8 $\pm$ 1.0      | 46.4             | 6.0                 |

Fits of single or double gaussian profiles to the histograms in Fig. 4 (5  $\mu$ M tilting/non-tilting and 0 ATP) and Supplementary Fig. 3 (40  $\mu$ M tilting/non-tilting) gave values for the peak axial angle ( $\bar{\beta}$ ) and standard deviation ( $\sigma_{\beta}$ ) of each gaussian profile. Uncertainties in these fitted parameters (for example,  $\pm 0.9$ ) are given as standard errors.  $\delta$ -value (wobble amplitude as described in the main text), mean ( $\bar{\delta}$ ) and standard deviation ( $\sigma_{\delta}$ ) were calculated for the population of steady  $\beta$ -levels in each gaussian profile (see Methods). Angles are given in degrees. *n* is the number of molecules. The number of steady  $\beta$ -levels in the 5- $\mu$ M-ATP tilting and 40- $\mu$ M-ATP tilting populations were 495 and 576, respectively.

and 13 s<sup>-1</sup>) that describe single-molecule mechanical stepping rates determined by optical trapping experiments<sup>11</sup>. Therefore, sequential tilts of the myosin V LCD each correspond to the binding of one ATP molecule, release of one ADP molecule, and a single mechanical step, which are all consistent with the hand-over-hand model.

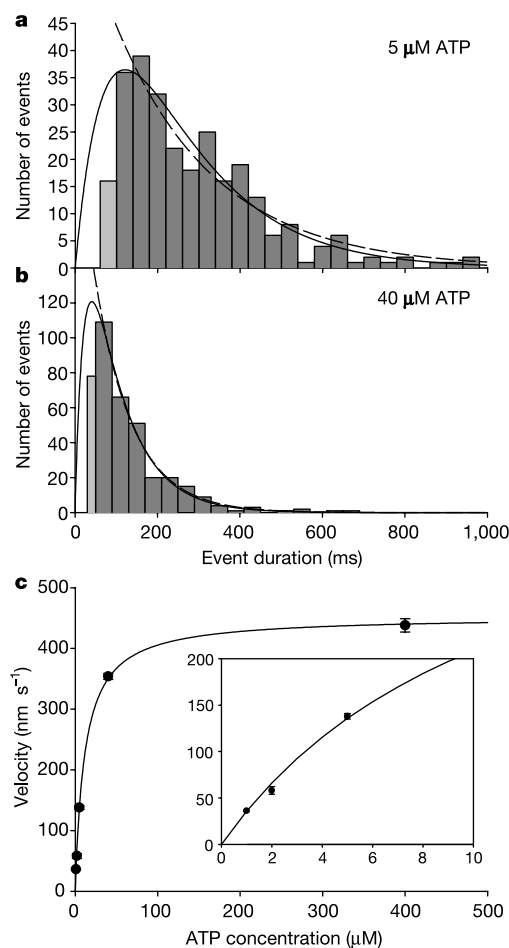
Average translocation velocities (*V*) were measured in the same experiments and were fitted by the hyperbolic relation  $V = V_{\max}[\text{ATP}]/([\text{ATP}] + K_M)$  (Fig. 5c), yielding  $V_{\max} = 453 \pm 10 \text{ nm s}^{-1}$  and  $K_M = 11.7 \pm 1.2 \mu\text{M}$  (95% confidence limits). Within uncertainties,  $K_M$  is equal to the ratio of  $k_2$  to  $k_1$ , implying that the same processes control the rate of the observed structural

changes and the mechanical displacements. To further quantify this relationship, the product of the velocity measured at each ATP concentration and the corresponding mean event duration,  $\tau_e = \left( \frac{1}{k_1[\text{ATP}]} + \frac{1}{k_2} \right)$ , was calculated to determine the average distance a myosin V molecule moves per discrete rotation of the LCD.  $V \cdot \tau_e$  values at 5 and 40  $\mu$ M ATP were  $36.2 \pm 4.0$  and  $37.4 \pm 3.5 \text{ nm}$  (95% confidence limits), respectively, similar to the crossover distance of actin's long-pitch helix. These values provide an estimate of the stride length of the myosin V molecule solely from its structural dynamics and independent of other mechanical or biochemical data. This stride length per structural change is comparable to the unitary mechanical displacements previously measured<sup>11</sup>, confirming the conclusion above that the LCD tilts once per mechanical step.

## Discussion

Our study represents the first direct time-resolved measurement of tilting in an actively translocating motor protein sufficient to explain the observed mechanical step. The head-tail junction of a myosin V molecule travels 36–37 nm for each major tilting motion, as shown above. Consider the lever arm provisionally to be a rigid 24-nm rod. The axis of the lever tilts, for instance, from 40° to 140° when stepping from the trailing to the leading position ((cos 40° – cos 140°)24 nm = 36.8 nm). With these angles, the head-tail junctions of the two heavy chains meet at the same radial distance from actin (sin 40° = sin 140°). Our rhodamine probe, oriented 40° from the lever axis (see Supplementary Information), is observed to tilt from about 30° to about 105° (Fig. 4, peak plotted near 75° = 180° – 105°). The slightly smaller magnitude of tilting can be accommodated if the probe dipole is not oriented in the plane of the lever arm rotation and the lever twists about its own axis by 10° or by 85° (calculation given in Supplementary Information). The probe data are compatible with other angles of lever tilt and twist if the CaM-binding region is flexible<sup>13</sup> or if the two head-tail junctions do not coincide. The data thus provide clear evidence for the lever arm function of the CaM-binding region of myosin V.

Single-molecule fluorescence polarization techniques have been used previously for studies of protein domain rotations<sup>20</sup>. Rotation of F<sub>1</sub>-ATPase<sup>21</sup> and actin<sup>22</sup>, and transitions between disordered and ordered states of myosin II<sup>23</sup> were detected during ATP-driven motility. Spontaneous motions of DNA<sup>24</sup> and kinesin<sup>25</sup> have been detected, the latter with a bifunctional probe similar to that used here. All of these previous studies were limited to detecting the two-dimensional projection of the probe orientation onto the plane of the microscope stage. With myosin V, determination of the lever arm rotations relative to actin required measurement of the three-dimensional probe orientation (that is, the probe angle relative to the microscope optical axis in addition to the orientation of the projection onto the stage). Other techniques have been described for measuring the three-dimensional orientations of single fluorophores<sup>20</sup>. However, these have not yet been used for biomolecular structural dynamics and are currently limited by specialized sample requirements<sup>26,27</sup> or low time resolution<sup>28,29</sup>. Robust, time-resolved measurement of three-dimensional probe orientation is a major advantage of the technique described here. Similar three-



**Figure 5** Event kinetics. **a**, **b**, Distributions of event dwell times determined from  $\beta$ -traces at 5  $\mu$ M (**a**) and 40  $\mu$ M (**b**) ATP (Fig. 1 and Supplementary Information). Dashed curves are single exponential fits; solid curves represent the global fit of a two-step model explained in the text. The first bins (light-grey bars) in **a** and **b** are under-populated owing to limited time resolution and were ignored during fitting. In **b**, the height of this bin was doubled because its width is halved. **c**, Average translocation velocities versus ATP concentration (inset shows an expanded view). Error bars indicate standard errors of the mean.

dimensional angular data should be relevant to many other biological systems that can operate near a surface.

A model of myosin V processivity must explain the following structural observations obtained in the present study: (1) in the presence of ATP, the LCD tilts between two discrete axial angles that differ by  $\geq 45^\circ$  or  $\geq 75^\circ$ ; (2) the two states corresponding to these orientations have approximately the same amplitude of 4 ns to 10 ms mobility; (3) the duration of occupancy in the two orientations is the same, within the 20–40-ms time resolution of the current experiments; (4) no additional structural states with distinct LCD orientations or mobilities are occupied for more than 20–40 ms; (5) myosin V translocates about 37 nm on average for each LCD rotation; and (6) the rates of the observed structural changes, combined with earlier measurements, imply one rotation per unitary mechanical step, and per hydrolysis of an ATP molecule.

These results imply that the approximately 37-nm mechanical steps of myosin V involve transitions of the individual heads between two structural states, both bound to actin strongly enough to limit the mobility of the CaM-binding region. These two states are probably a leading head at the beginning of its working stroke and a trailing head at the end. Several models other than the hand-over-hand mechanism have been proposed to explain processive translocation of motor proteins along cytoskeletal tracks—in particular, ‘biased diffusion’, ‘inchworm’ and ‘hotspot’ models<sup>30–35</sup>. Biased diffusion and hotspot models, as well as some hand-over-hand models<sup>36</sup> postulate that, for a significant portion of a catalytic cycle, only one head of the motor protein is strongly bound to its track. They typically do not include high occupancy of the state with two heads firmly attached to actin that predominates in the present experiments. In the typical inchworm model<sup>33</sup>, the structural state of neither head changes on completion of consecutive steps, which is incompatible with our observation that individual heads oscillate between two distinct orientations. Of the proposed models for myosin V processivity, the hand-over-hand model, including a working stroke generated by a tilting lever arm, most readily explains the observations reported here.

*Note added in proof:* Burgess *et al.*<sup>44</sup> have recently published results that are difficult to reconcile with one of our possible explanations for the non-tilting CaM light chains in the presence of ATP: the bent neck model. □

## Methods

### Protein preparation

Rabbit muscle actin was purified<sup>37</sup> and labelled<sup>38</sup> at Cys 374 with 1,5-IAEDANS. F-actin, at 1  $\mu$ M total monomer concentration, 14% unlabelled, 81% AEDANS-labelled, and 5% biotin-labelled (Cytoskeleton), was stabilized with 2.6  $\mu$ M phalloidin (Molecular Probes). Residues Pro 66 and Ala 73 of chicken CaM<sup>39</sup> were mutated to Cys (QuikChange, Stratagene; plasmid supplied by H. L. Sweeney). Mutant CaM was expressed<sup>40</sup>, labelled<sup>17</sup> and purified (Supplementary Information). Electrospray mass spectrometry, after endoproteinase Asp-N (Roche Diagnostics) digestion, confirmed rhodamine cross-linking of Cys 66 and Cys 73. Chick brain myosin V was purified<sup>41</sup> and labelled by exchanging<sup>12</sup> endogenous CaM with a 1:10 mixture of bifunctional rhodamine-labelled calmodulin and wild-type CaM.

### Motility assay

Fused silica slides (Quartz Scientific), spin-coated with 2 mg ml<sup>−1</sup> poly(methyl methacrylate) (Aldrich Chemical, no. 37,003-7) in methylene chloride, were incorporated into approximately 10- $\mu$ l sample chambers with glass coverslips and double-sided adhesive tape. F-actin was bound to the poly(methyl methacrylate) surface through biotinylated BSA and streptavidin<sup>12</sup>. Flow aligned the F-actin with the microscope x axis. Myosin V labelled with bifunctional rhodamine-labelled calmodulin was introduced into the sample chamber at about 10–1,000 pM in 25 mM KCl, 2 mM MgCl<sub>2</sub>, 1 mM EGTA, 100  $\mu$ g ml<sup>−1</sup> wild-type CaM, 100 mM dithiothreitol, 20 mM HEPES, pH 7.6, and ATP as indicated. Compatible with earlier work<sup>8</sup>,  $k_1$  and  $K_M$  were calculated based on total ATP concentration. With the dissociation constant ( $K_d$ ) for Mg<sup>2+</sup> binding to ATP of 133  $\mu$ M, MgATP concentration would be 6% less<sup>42</sup>. This correction does not affect calculated step sizes. For 0 ATP experiments, the buffer included 2 U ml<sup>−1</sup> apyrase (Sigma).

### Single-molecule fluorescence polarization

Full details of this technique will be described elsewhere (J.N.F., M.E.Q. and Y.E.G., manuscript in preparation). Briefly, a 1 mW, 355-nm laser (JDS Uniphase) and a 10–20 mW, 532-nm laser (Lightwave Electronics) were directed towards the sample

chamber through a BK7 prism (ESCO Products) and index-matching liquid (Cargille Laboratories) at a glancing angle relative to the microscope slide. Total internal reflection generated evanescent waves near the sample chamber surface<sup>18,19</sup>. The 532-nm laser direction and polarization were modulated by Pockels cells (Conoptics) at 5 or 10 ms, producing four excitation polarizations. s1 and s2 generated linear polarizations along y and x axes; p1 and p2 produced polarizations approximately along the z axis with small components along x and y, respectively<sup>20</sup>. Camera (V/ICCD, Roper Scientific) images were used to identify actin filament location and orientation, and individual rhodamines. A piezo-electric stage (Polytec PI) centred a single labelled myosin V molecule co-localized with actin on the microscope stage. A  $\times 100$  1.2 NA water immersion objective projected fluorescence through a polarizing beam splitter to resolve x and y polarizations onto two avalanche photodiodes (Perkin Elmer, SPCM-AQR-16). Total fluorescence was calculated using  $I_{\text{Tot}} = (s_1 I_x + s_1 I_y + s_2 I_x + s_2 I_y) + 0.885(p_1 I_x + p_1 I_y + p_2 I_x + p_2 I_y)$ , which is approximately independent of probe orientation.

### Data analysis

Single molecules were identified by total intensity and single-step photobleaching. In traces with two bleaching steps, only data after the first step were analysed. Analytical equations (to be described elsewhere: J.N.F., M.E.Q. and Y.E.G., manuscript in preparation) were fitted to data for each time point to obtain  $I_{\text{Fit}}$ ,  $\beta$ ,  $\alpha$  and  $\delta$ . Transitions where the  $\beta$ -value changed abruptly from one approximately constant value to another were identified manually using defined criteria. For averaging, transitions were grouped according to whether the  $\beta$ -value increased or decreased. Each transition was shifted to set the preceding datum to 0 ms and traces were averaged between the preceding and following adjacent transitions.

### Histograms of $\beta$ -values

$\beta$ -values were averaged for each region of a trace where  $\beta$  remained approximately constant (steady  $\beta$ -level), except the last region in traces exhibiting tilting. Three-degree bin width histograms of averages were fitted by gaussian distributions,  $\rho(\beta)$ , given by:

$$\rho(\beta) = A \left( \frac{f}{\sigma_{\beta 1} \sqrt{2\pi}} \right) \exp \left[ -\frac{1}{2} \left( \frac{\beta - \bar{\beta}_1}{\sigma_{\beta 1}} \right)^2 \right] + A \left( \frac{1-f}{\sigma_{\beta 2} \sqrt{2\pi}} \right) \exp \left[ -\frac{1}{2} \left( \frac{\beta - \bar{\beta}_2}{\sigma_{\beta 2}} \right)^2 \right]$$

where  $A$  is the product of the total number of steady  $\beta$ -levels and the bin width;  $f$  is the fraction of steady  $\beta$ -levels in the first gaussian peak. For 5 and 40  $\mu$ M ATP data exhibiting tilting, mean  $\pm$  s.d. of  $\delta$ -values were determined for the two groups of steady  $\beta$ -intervals with average  $\beta$  greater or less than  $50^\circ$ . For 0 ATP data, the groups were split at  $39^\circ$ . For 5 and 40  $\mu$ M ATP data without tilting, they were determined for average  $\beta$  within  $\bar{\beta}_2 \pm 3\sigma_{\beta 2}$ .

### Event dwell times

Dwell times were differences between the starting times of consecutive events with  $\beta$ -values on opposite sides of  $50^\circ$ . Separating events into populations with  $\beta$ -values greater or less than  $50^\circ$  gave single exponential fits to histograms of 5  $\mu$ M ATP dwell times, with rate constants of  $4.2 \pm 0.3$  s<sup>−1</sup> and  $4.1 \pm 0.3$  s<sup>−1</sup> ( $\pm$  standard error), respectively. The corresponding rates for the 40  $\mu$ M ATP data were  $10.1 \pm 0.4$  s<sup>−1</sup> and  $11.7 \pm 0.7$  s<sup>−1</sup>. A single exponential fit to histograms with all event data (without discriminating by  $\beta$ -value) gave rate constants of  $4.1 \pm 0.2$  s<sup>−1</sup> for 5  $\mu$ M ATP and  $10.9 \pm 0.4$  s<sup>−1</sup> for 40  $\mu$ M ATP.

The 5 and 40  $\mu$ M ATP histograms were also fit simultaneously to:

$$\rho(t) = A \left( \frac{k_1 [\text{ATP}] k_2}{k_1 [\text{ATP}] - k_2} \right) (\exp(-k_2 t) - \exp(-k_1 [\text{ATP}] t))$$

where  $t$  is dwell time and  $A$  is amplitude (different for the two ATP concentrations). This fit used bin widths equal to the measurement time resolutions (MTR): 40 ms for 5  $\mu$ M ATP and 20 ms for 40  $\mu$ M ATP. Because selection criteria required events to contain  $\geq 2$  time points, no dwell times equal to the MTR were scored. Bins at twice the MTR also had fewer events than expected from the fitted curves, presumably being undercounted owing to noise. These bins (light-grey bars in Fig. 5) were therefore ignored. Monte-Carlo simulations determined 95% confidence limits for fitted rates<sup>43</sup>.

### Velocity measurements

Velocities of individual myosin V molecules were determined from images with 10-ms-resolution time stamps taken at approximately 1-s intervals.

Received 16 December 2002; accepted 3 March 2003; doi:10.1038/nature01529.

- Reedy, M. K., Holmes, K. C. & Tregear, R. T. Induced changes in orientation of the cross-bridges of glycerinated insect flight muscle. *Nature* **207**, 1276–1280 (1965).
- Rayment, I. *et al.* Structure of the actin–myosin complex and its implications for muscle contraction. *Science* **261**, 58–65 (1993).
- Spudich, J. A. The myosin swinging cross-bridge model. *Nature Rev. Mol. Cell Biol.* **2**, 387–392 (2001).
- Geeves, M. A. & Holmes, K. C. Structural mechanism of muscle contraction. *Annu. Rev. Biochem.* **68**, 687–728 (1999).
- Irving, M. & Piazzesi, G. Motions of myosin heads that drive muscle contraction. *News Physiol. Sci.* **12**, 249–254 (1997).
- Goldman, Y. E. Wag the tail: Structural dynamics of actomyosin. *Cell* **93**, 1–4 (1998).
- Cheney, R. E. *et al.* Brain myosin-V is a two-headed unconventional myosin with motor activity. *Cell* **75**, 13–23 (1993).
- De La Cruz, E. M., Wells, A. L., Rosenfeld, S. S., Ostap, E. M. & Sweeney, H. L. The kinetic mechanism of myosin V. *Proc. Natl Acad. Sci. USA* **96**, 13726–13731 (1999).
- Mehta, A. D. *et al.* Myosin-V is a processive actin-based motor. *Nature* **400**, 590–593 (1999).
- Mehta, A. Myosin learns to walk. *J. Cell Sci.* **114**, 1981–1998 (2001).
- Rief, M. *et al.* Myosin-V stepping kinetics: A molecular model for processivity. *Proc. Natl Acad. Sci. USA* **97**, 9482–9486 (2000).

12. Sakamoto, T., Amitani, I., Yokota, E. & Ando, T. Direct observation of processive movement by individual myosin V molecules. *Biochem. Biophys. Res. Commun.* **272**, 586–590 (2000).
13. Walker, M. L. *et al.* Two-headed binding of a processive myosin to F-actin. *Nature* **405**, 804–807 (2000).
14. Veigel, C., Wang, F., Bartoo, M. L., Sellers, J. R. & Molloy, J. E. The gated gait of the processive molecular motor, myosin V. *Nature Cell Biol.* **4**, 59–65 (2002).
15. Moore, J. R., Kremmentsova, E. B., Trybus, K. M. & Warshaw, D. M. Myosin V exhibits a high duty cycle and large unitary displacement. *J. Cell Biol.* **155**, 625–635 (2001).
16. Corrie, J. E. T. *et al.* Dynamic measurement of myosin light-chain-domain tilt and twist in muscle contraction. *Nature* **400**, 425–430 (1999).
17. Corrie, J. E. T., Craik, J. S. & Munasinghe, V. R. N. A homobifunctional rhodamine for labeling proteins with defined orientations of a fluorophore. *Bioconjug. Chem.* **9**, 160–167 (1998).
18. Axelrod, D., Burghardt, T. P. & Thompson, N. L. Total internal reflection fluorescence. *Annu. Rev. Biophys. Bioeng.* **13**, 247–268 (1984).
19. Funatsu, T., Harada, Y., Tokunaga, M., Saito, K. & Yanagida, T. Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution. *Nature* **374**, 555–559 (1995).
20. Forkey, J. N., Quinlan, M. E. & Goldman, Y. E. Protein structural dynamics by single-molecule fluorescence polarization. *Prog. Biophys. Mol. Biol.* **74**, 1–35 (2000).
21. Adachi, K. *et al.* Stepping rotation of F<sub>1</sub>-ATPase visualized through angle-resolved single-fluorophore imaging. *Proc. Natl Acad. Sci. USA* **97**, 7243–7247 (2000).
22. Sase, I., Miyata, H., Ishiwata, S. & Kinosita, K. Jr Axial rotation of sliding actin filaments revealed by single-fluorophore imaging. *Proc. Natl Acad. Sci. USA* **94**, 5646–5650 (1997).
23. Warshaw, D. M. *et al.* Myosin conformational states determined by single fluorophore polarization. *Proc. Natl Acad. Sci. USA* **95**, 8034–8039 (1998).
24. Ha, T., Glass, J., Enderle, Th., Chemla, D. S. & Weiss, S. Hindered rotational diffusion and rotational jumps of single molecules. *Phys. Rev. Lett.* **80**, 2093–2096 (1998).
25. Sosa, H., Peterman, E. J. G., Moerner, W. E. & Goldstein, L. S. B. ADP-induced rocking of the kinesin motor domain revealed by single-molecule fluorescence polarization microscopy. *Nature Struct. Biol.* **8**, 540–544 (2001).
26. Bopp, M. A., Jia, Y., Li, L., Cogdell, R. J. & Hochstrasser, R. M. Fluorescence and photobleaching dynamics of single light-harvesting complexes. *Proc. Natl Acad. Sci. USA* **94**, 10630–10635 (1997).
27. Empedocles, S. A., Neuhauser, R. & Bawendi, M. G. Three-dimensional orientation measurements of symmetric single chromophores using polarization microscopy. *Nature* **399**, 126–130 (1999).
28. Betzig, E. & Chichester, R. J. Single molecules observed by near-field scanning optical microscopy. *Science* **262**, 1422–1425 (1993).
29. Dickson, R. M., Norris, D. J. & Moerner, W. E. Simultaneous imaging of individual molecules aligned both parallel and perpendicular to the optic axis. *Phys. Rev. Lett.* **81**, 5322–5325 (1998).
30. Vale, R. D. & Oosawa, F. Protein motors and Maxwell's demons: Does mechanochemical transduction involve a thermal ratchet? *Adv. Biophys.* **26**, 97–134 (1990).
31. Block, S. M. Kinesin: What gives? *Cell* **93**, 5–8 (1998).
32. Okada, Y. & Hirokawa, N. A processive single-headed motor: Kinesin superfamily protein KIF1A. *Science* **283**, 1152–1157 (1999).
33. Hua, W., Chung, J. & Gelles, J. Distinguishing inchworm and hand-over-hand processive kinesin movement by neck rotation measurements. *Science* **295**, 844–848 (2002).
34. Tanaka, H. *et al.* The motor domain determines the large step of myosin-V. *Nature* **415**, 192–195 (2002).
35. Nishikawa, S. *et al.* Class VI myosin moves processively along actin filaments backward with large steps. *Biochem. Biophys. Res. Commun.* **290**, 311–317 (2002).
36. De La Cruz, E. M., Ostap, E. M. & Sweeney, H. L. Kinetic mechanism and regulation of myosin VI. *J. Biol. Chem.* **276**, 32373–32381 (2001).
37. Pardee, J. D. & Spudich, J. A. Purification of muscle actin. *Methods Cell Biol.* **24**, 271–289 (1982).
38. Cooper, J. A., Walker, S. B. & Pollard, T. D. Pyrene actin: Documentation of the validity of a sensitive assay for actin polymerization. *J. Muscle Res. Cell Motil.* **4**, 253–262 (1983).
39. Putkey, J. A. *et al.* Chicken calmodulin genes. *J. Biol. Chem.* **258**, 11864–11870 (1983).
40. Putkey, J. A., Slaughter, G. R. & Means, A. R. Bacterial expression and characterization of proteins derived from the chicken calmodulin cDNA and a calmodulin processed gene. *J. Biol. Chem.* **260**, 4704–4712 (1985).
41. Cheney, R. E. Purification and assay of myosin V. *Methods Enzymol.* **298**, 3–18 (1998).
42. Moiescu, D. G. & Thieleczek, R. Sarcomere length effects on the Sr<sup>2+</sup>- and Ca<sup>2+</sup>-activation curves in skinned frog muscle fibres. *Biochim. Biophys. Acta* **546**, 64–76 (1979).
43. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in C: The Art of Scientific Computing* 2nd edn (Cambridge Univ. Press, Cambridge, UK, 1992).
44. Burgess, S. *et al.* The prepower stroke conformation of myosin V. *J. Cell. Biol.* **159**, 983–991 (2002).

**Supplementary Information** accompanies the paper on Nature's website  
(<http://www.nature.com/nature>).

**Acknowledgements** This work was supported by the NIH and Medical Research Council. We thank H. Shuman, M. Ostap, H. Higuchi, M. Mooseker, R. Cheney, A. Houdusse, D. Trentham, R. Ferguson, B. Brandmeier and F. Vanzi for discussions, and R. Munasinghe, I. Gertsman, S. Manz, D. Keleti and N. Tang for technical assistance.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to Y.E.G. (e-mail: goldmany@mail.med.upenn.edu).



# An energetic stellar outburst accompanied by circumstellar light echoes

Howard E. Bond\*, Arne Henden†, Zoltan G. Levay\*, Nino Panagia\*‡, William B. Sparks\*, Sumner Starrfield§, R. Mark Wagner||, R. L. M. Corradi¶ & U. Munari#

\* Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA

† Universities Space Research Association & US Naval Observatory, Flagstaff Station, PO Box 1149, Flagstaff, Arizona 86002, USA

§ Department of Physics & Astronomy, Arizona State University, Tempe, Arizona 85287-1504, USA

|| Large Binocular Telescope Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, Arizona 85721, USA

¶ Isaac Newton Group of Telescopes, Apartado de Correos 321, 38700 Santa Cruz de La Palma, Canarias, Spain

# INAF-Osservatorio Astronomico di Padova, Sede di Asiago, 36012 Asiago (VI), Italy

Some classes of stars, including novae and supernovae, undergo explosive outbursts that eject stellar material into space. In 2002, the previously unknown variable star V838 Monocerotis brightened suddenly by a factor of  $\sim 10^4$ . Unlike a supernova or nova, it did not explosively eject its outer layers; rather, it simply expanded to become a cool supergiant with a moderate-velocity stellar wind. Superluminal light echoes were discovered<sup>1,2</sup> as light from the outburst propagated into the surrounding, pre-existing circumstellar dust. Here we report high-resolution imaging and polarimetry of those light echoes, which allow us to set direct geometric distance limits to the object. At a distance of  $>6$  kpc, V838 Mon at its maximum brightness was temporarily the brightest star in the Milky Way. The presence of the circumstellar dust implies that previous eruptions have occurred, and spectra show it to be a binary system. When combined with the high luminosity and unusual outburst behaviour, these characteristics indicate that V838 Mon represents a hitherto unknown type of stellar outburst, for which we have no completely satisfactory physical explanation.

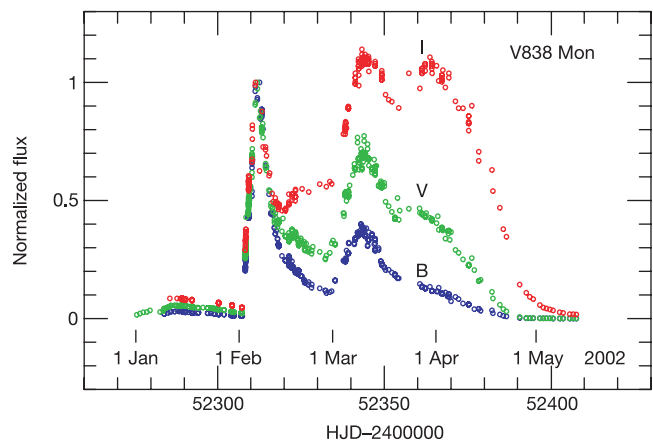
Light echoes around a star undergoing an outburst appear after the eruption—the illumination is delayed by the longer path length from the star to the surrounding dust and then to the Earth, as compared with the light from the outburst itself travelling directly to the Earth. In the case of an instantaneous light pulse, the geometry of a light echo is straightforward<sup>3–8</sup>: the illuminated dust lies on the paraboloid given by  $z = x^2/2ct - ct/2$ , where  $x$  is the projected distance from the star in the plane of the sky,  $z$  is the distance from this plane along the line of sight towards the Earth,  $c$  is the speed of light, and  $t$  is the time since the outburst. Light echoes are rare, having been observed only around a few classical novae in our own Galaxy<sup>3,9,10</sup>, and around several extragalactic supernovae<sup>11</sup>. With the exception of SN1987A (ref. 12), these echoes arose from interstellar dust or gas in the line of sight, rather than from true circumstellar material.

V838 Mon erupted in early January 2002 (ref. 13). Figure 1 shows the light curve of the outburst in linear flux units, from ground-based observations in three different wavelength bands: blue (B), visual (V) and near-infrared (I). This light curve—with a weak precursor brightening, a sharp, bright blue peak, followed by a decline and then two further broad and much redder peaks—is unlike that of a supernova, nova, or any other known type of variable star. Spectra reported by other workers (data not shown)

showed a very cool photosphere throughout the outburst. Toward the end of the outburst, and during the decline, the star became extremely red. Recent spectra<sup>14–16</sup>, taken after the decline, show that a hot B-type star is also present, along with the cool outbursting object.

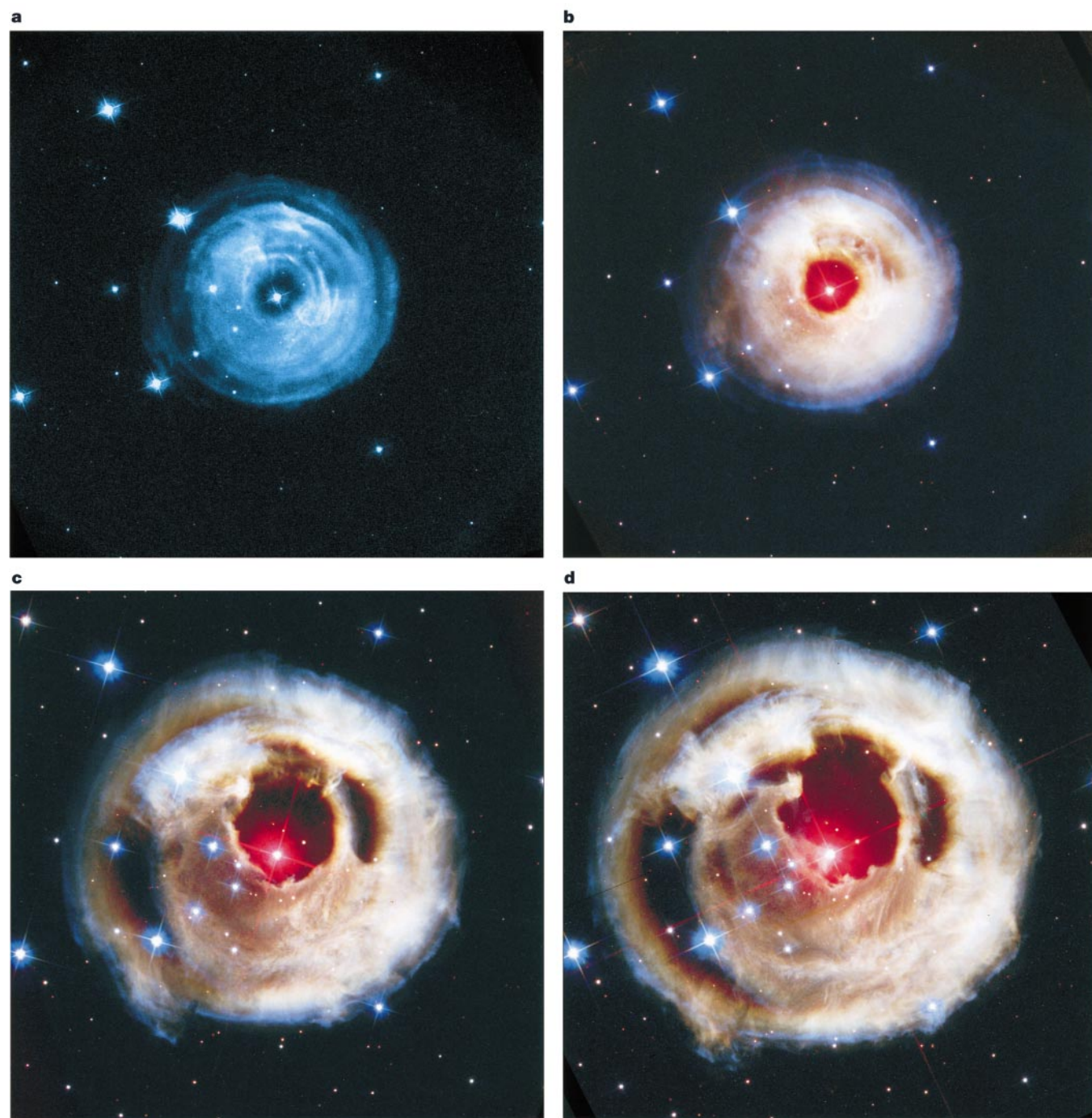
We have obtained Hubble Space Telescope (HST) observations of the light echoes surrounding V838 Mon with the Advanced Camera for Surveys (ACS) at the four epochs listed in Table 1. Figure 2 shows the ACS images. Only the B filter was used at the first epoch, but for the others we obtained three-colour (BVI) frames, and have used these to make renditions that approximate true colour. The nebula exhibits a wealth of subarcsecond cirrus-like structure, but is dominated (especially at the first two epochs) by a remarkable series of nearly circular arcs and rings, centred on the variable star. The exact linear size of the illuminated dust envelope depends on the distance to the object, but on the basis of the distance limits given below, the visible material is confined to less than 2 pc from the star and is thus circumstellar, rather than being ambient interstellar medium. The cavity centred on the star (Fig. 2a, b), along with several features that appear aligned towards the star, also establish the circumstellar nature of the echoes.

As shown in Fig. 1, the outburst light curve had the simplest structure in the B band—a bright, relatively short peak on 6 February, and a second, weaker peak about 32 days later—making the B-band light echoes most amenable to geometric analysis. If we assume the simplest geometry, which is a series of nested spherical dust shells centred on the star and intersecting the expanding light paraboloids, then it can be shown that the apparent expansion rate of a ring is given by  $dx/dt = (rc - c^2t)(2rct - c^2t^2)^{-1/2}$ , where  $r$  is the radius of the ring measured from the star. Thus the time behaviour of these rings in angular units provides a direct geometric distance to the object. The pair of images taken only 20.8 days apart,



**Figure 1** Multicolour light curves of the outburst of V838 Mon. Data consist of observations by A.H. (at the US Naval Observatory, Flagstaff) and by amateurs (<http://vsnet.kusastro.kyoto-u.ac.jp/vsnet/index.html>) and by other professional observers<sup>2,17,26–29</sup>. Magnitudes in the standard Johnson blue (B), visual (V), and Kron-Cousins near-infrared (I) bands have been converted to flux on a linear scale, normalized to the peak at heliocentric Julian date (HJD) 2452311. Corresponding calendar dates are shown above. After an initial rapid rise to tenth magnitude and a subsequent gradual decline, V838 Mon abruptly rose another four magnitudes in early February, reaching a maximum of  $V = 6.75$  mag on 6 February 2002. After another decline, the star rose to a second peak in early March. Once again there was a slow decline after this peak, except in the I band, where yet a third peak was attained in early April. In late April, V838 Mon dropped quickly back to its original quiescent brightness in the B and V bands; this precipitous drop in brightness has been attributed to dust formation above the photosphere<sup>29</sup>. The progenitor object was recorded in archival sky surveys<sup>2</sup> at  $V = 15.6$  mag, very similar to its current post-outburst brightness.

‡ On assignment from the European Space Agency, Space Telescope Division.



**Figure 2** HST images of the light echoes. The apparently superluminal expansion of the echoes as light from the outburst propagates outward into surrounding dust is shown dramatically. Images were taken in 2002 on 30 April (**a**), 20 May (**b**), 2 September (**c**) and 28 October (**d**). Each frame is  $83'' \times 83''$ ; north is up and east to the left. Imaging on 30 April was obtained only in the B filter, but B, V and I were used on the other three dates, allowing us to make full-colour renditions. The time evolution of the stellar outburst (Fig. 1)

on 30 April and 20 May, is the most suitable for such a measurement, as the same rings can be identified unambiguously in both images. The expansion rates are plotted against the mean angular radii of each ring in Fig. 3. Overlaid on the plot are the expansion rates calculated from the above equation, for ages of 62 and 93 days, which are the elapsed times between the first and second B-band light-curve peaks and the mean epoch of the two images. It can be seen that the measured points fall into two families, plotted for

is reflected by structures visible in these colour images. In **b**, for example, note the series of rings and filamentary structures, especially in the upper right quadrant. Close examination shows that each set of rings has a sharp, blue outer edge, a dip in intensity nearer the star, and then a rebrightening to a redder plateau. Similar replicas of the outburst light curve are seen propagating outwards throughout all of the colour images.

clarity as filled and open circles. The filled circles satisfactorily follow the expected behaviour if the corresponding echo rings are illuminated by the first light-curve peak, while the open circles appear to represent rings illuminated by the second peak. In strong confirmation of this interpretation, we find that the filled circles all refer to the blue rings seen clearly in Fig. 2b, while the open circles all correspond to the trailing, red-edged rings.

The data in Fig. 3 show significant scatter, probably arising from



Table 1 HST observations

| UT date<br>(2002) | HJD       | Days since<br>Feb. 6.4 | Days since<br>Mar. 10.0 | ACS filters used    |
|-------------------|-----------|------------------------|-------------------------|---------------------|
| Apr. 30           | 2452394.7 | 82.8                   | 51.2                    | B + pol             |
| May 20            | 2452415.5 | 103.6                  | 72.0                    | B + pol, V + pol, I |
| Sep. 2            | 2452520.4 | 208.5                  | 176.9                   | B, V + pol, I       |
| Oct. 28           | 2452576.0 | 264.1                  | 232.5                   | B, V, I             |

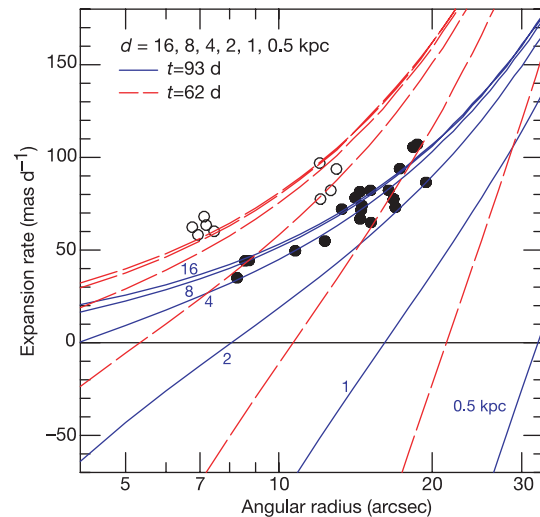
The ACS filters F435W, F606W and F814W are denoted B, V and I, respectively, and ' + pol' denotes that the indicated filter was combined in succession with 0°, 60° and 120° polarizers. The time of maximum B luminosity was near 2002 February 6.4, and the second B peak was near March 10.0. (The lack of HST observations between May and September is due to solar avoidance.)

departures of the actual circumstellar structures from the simple adopted spherical geometry. For plausible non-spherical geometries, measures made over several azimuths will still average to a result close to the true distance. Thus, barring a very unusual morphology, the figure implies that the distance to V838 Mon must certainly be greater than 2 kpc, and probably significantly greater. As the predicted expansion rates approach an asymptotic limit with increasing distance, we cannot make a stronger statement at present. However, because the relations for various distances diverge with increasing time, continued high-resolution imaging will eventually yield a very precise distance.

A second, independent geometrical method for determining the distance comes from polarimetry<sup>8</sup>. Because maximum linear polarization occurs for 90° scattering, there will be highly polarized light at a linear radius  $ct$ , whose angular radius would thus yield the distance—but of course only if there actually is dust at that radius and at the same distance from us as the star. Our most recent polarimetric HST image, obtained on 2 September 2002, shows very high polarization (~50%), but only at the inner rim of the cavity at the lower left (southeast) of the star. Thus, at present, we can only set a lower limit to the distance, by equating the radius at this rim, ~5.0", to  $ct$ , where  $t$  is now the time since the second light-curve peak. This lower limit is ~6 kpc. Future observations will provide a definitive distance from this method as well, once the light-echo paraboloid enlarges enough to allow illumination of close-in material behind, as well as in front of, the star. At that time, we expect to see maximum polarization in a ring<sup>8</sup>, rather than right at the inner edge of the rim.

For a distance of at least 6 kpc (and interstellar reddening of  $E(B - V) \approx 0.8$ ; refs 17, 18), the absolute magnitude of V838 Mon at maximum was at least  $M_V = -9.6$ , making it more luminous than a classical nova, and temporarily the brightest star in the Milky Way. The light curve, spectroscopic behaviour, and high luminosity of V838 Mon are reminiscent of "M31 RV," a luminous red variable star that appeared in the Andromeda galaxy in the late 1980s (ref. 19). M31 RV reached an absolute visual magnitude at least as bright as  $M_V = -9.3$  (ref. 20), but unfortunately was not well studied before it declined below detectability. V838 Mon and M31 RV (along with, possibly, the Galactic object V4332 Sagittarii; ref. 21) appear to represent a new type of outburst in which a star expands rapidly to supergiant dimensions<sup>22</sup>, but without either the catastrophic envelope loss or the evolution to high temperature seen in classical and symbiotic novae.

In the case of V838 Mon the outburst appears to have occurred in a binary system, which may suggest an analogy with a nova system in which a thermonuclear explosion occurs on an accreting white-dwarf companion. However, a broad theoretical exploration<sup>23</sup> of the parameter space for thermonuclear runaways on accreting white dwarfs predicts no behaviour like that of V838 Mon. Evidently, V838 Mon has in the recent past had several similar outbursts, which produced the dust envelope that is now being illuminated. These previous outbursts appear to rule out one-time catastrophic events such as a stellar collision or merger<sup>24</sup>, or a common-envelope



**Figure 3** Measured and predicted apparent angular expansion rates of rings seen in the light echoes. Measurements were done on the pair of B-band images taken on 30 April and 20 May 2002. A simple edge-finding algorithm was used to mark the locations of the same rings on both images at several different azimuths, and their angular expansion rates were calculated. Filled circles denote rings that we attribute to illumination from the first light-curve peak, and open circles denote rings attributed to the second peak. Measurement errors are only slightly larger than the size of the plotted points. Units of expansion are milliarcseconds per day, as functions of angular radius in arcseconds. Also shown are the predicted expansion rates for light-echo rings seen 62 days (red dashed lines) and 93 days (blue solid lines) after outburst, calculated from the equation given in the text. They are functions of distance to the star, plotted for the range from 0.5 to 16 kpc, and are apparently superluminal owing to the parabolic geometry of the illuminated surfaces. The measured angular expansion rates imply a strong lower limit to the stellar distance of 2 kpc.

interaction<sup>25</sup>, as the energy source. Now that V838 Mon seems to be returning to its quiescent state, further observations of the system may shed light on its properties and give insight into possible outburst mechanisms for this extraordinary object. □

Received 26 November 2002; accepted 18 February 2003; doi:10.1038/nature01508.

- Henden, A. *et al.* V838 Monocerotis. *IAU Circ. No. 7859* (2002).
- Munari, U. *et al.* The mysterious eruption of V838 Mon. *Astron. Astrophys.* **389**, L51–L56 (2002).
- Schaefer, B. E. Light echoes: novae. *Astrophys. J.* **327**, 347–349 (1988).
- Kapteyn, J. C. Über die Bewegung der Nebel in der Umgebung von Nova Persei. *Astron. Nachr.* **157**, 201–204 (1902).
- Couderc, P. Les auroles lumineuses des novae. *Ann. Astrophys.* **2**, 271–302 (1939).
- Russell, H. N., Dugan, R. S. & Stewart, J. Q. *Astronomy: a Revision of Young's Manual of Astronomy* 786 (Ginn, Boston, 1945).
- Chevalier, R. A. The scattered-light echo of a supernova. *Astrophys. J.* **308**, 225–231 (1986).
- Sparks, W. B. A direct way to measure the distances of galaxies. *Astrophys. J.* **433**, 19–28 (1994).
- Ritchey, G. W. Nebulosity about Nova Persei. Recent photographs. *Astrophys. J.* **15**, 129–131 (1902).
- Casalegno, R. *et al.* The emission nebula associated with V1974 Cygni: a unique object? *Astron. Astrophys.* **361**, 725–733 (2000).
- Sparks, W. B. *et al.* Evolution of the light echo of SN 1991T. *Astrophys. J.* **523**, 585–592 (1999).
- Bond, H. E., Gilmozzi, R., Meakes, M. G. & Panagia, N. Discovery of an inner light-echo ring around SN 1987A. *Astrophys. J.* **354**, L49–L52 (1990).
- Brown, N. J. Peculiar variable in Monoceros. *IAU Circ. No. 7785* (2002).
- Geballe, T. R., Smalley, B., Evans, A. & Rushton, M. T. V838 Monocerotis. *IAU Circ. No. 8016* (2002).
- Munari, U. & Desidera, S. V838 Monocerotis. *IAU Circ. No. 8005* (2002).
- Wagner, R. M. & Starrfield, S. V838 Monocerotis. *IAU Circ. No. 7992* (2002).
- Kimeswenger, S., Lederle, C., Schmeja, S. & Armsdorfer, B. The peculiar variable V838 Mon. *Mon. Not. R. Astron. Soc.* **336**, L43–L47 (2002).
- Zwitter, T. & Munari, U. V838 Monocerotis. *IAU Circ. No. 7812* (2002).
- Mould, J. *et al.* A nova-like red variable in M31. *Astrophys. J.* **353**, L35–L37 (1990).
- Bryan, J. & Royer, R. E. Photometry of the unique luminous red variable in M31. *Publ. Astron. Soc. Pacif.* **104**, 179–181 (1992).
- Martini, P. *et al.* Nova Sagittarii 1994 I (V4332 Sagittarii): The discovery and evolution of an unusual luminous red variable star. *Astron. J.* **118**, 1034–1042 (1999).
- Munari, U., Henden, A., Corradi, R. L. M. & Zwitter, T. in *Classical Nova Explosions* (eds Hernanz, M. & Jose, J.) 52–56 (American Institute of Physics, New York, 2002).



23. Prialnik, D. & Kovetz, A. An extended grid of multicycle nova evolution models. *Astrophys. J.* **445**, 789–810 (1995).

24. Soker, N. & Tylenda, R. Main-sequence stellar eruption model for V838 Monocerotis. *Astrophys. J.* **582**, L105–L108 (2003).

25. Iben, I. & Tutukov, A. V. Rare thermonuclear explosions in short-period cataclysmic variables, with possible application to the nova-like red variable in the Galaxy M31. *Astrophys. J.* **389**, 369–374 (1992).

26. Goranskii, V. P. *et al.* Nova Monocerotis 2002 (V838 Mon) at early outburst stages. *Astron. Lett.* **28**, 691–700 (2002).

27. Osiwala, J. P. *et al.* The double outburst of the unique object V838 Mon. in *Symbiotic Stars Probing Stellar Evolution* (eds Corradi, R. L. M., Mikolajewska, J. & Mahoney, T. J.) (ASP Conference Series, San Francisco, in the press).

28. Price, A. *et al.* Multicolor observations of V838 Mon. *IAU Inform. Bull. Var. Stars* **5315**, 1 (2002).

29. Crause, L. A. *et al.* The post-outburst photometric behaviour of V838 Mon. *Mon. Not. R. Astron. Soc.* (in the press).

**Acknowledgements** This Letter is based on observations with the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute, which is operated by AURA, Inc. under a NASA contract. We thank the Space Telescope Science Institute for awarding Director's Discretionary Observing Time for this project, and for support. S.S. was supported, in part, by NSF and NASA grants to Arizona State University.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to H.E.B. (e-mail: bond@stsci.edu).

Realization of the Cirac–Zoller controlled-NOT quantum gate

Ferdinand Schmidt-Kaler, Hartmut Häffner, Mark Riebe, Stephan Gulde, Gavin P. T. Lancaster, Thomas Deuschle, Christoph Becher, Christian F. Roos, Jürgen Eschner & Rainer Blatt

Institut für Experimentalphysik, Universität Innsbruck, Technikerstraße 25, A-6020 Innsbruck, Austria

Quantum computers have the potential to perform certain computational tasks more efficiently than their classical counterparts. The Cirac–Zoller proposal<sup>1</sup> for a scalable quantum computer is based on a string of trapped ions whose electronic states represent the quantum bits of information (or qubits). In this scheme, quantum logical gates involving any subset of ions are realized by coupling the ions through their collective quantized motion. The main experimental step towards realizing the scheme is to implement the controlled-NOT (CNOT) gate operation between two individual ions. The CNOT quantum logical gate corresponds to the XOR gate operation of classical logic that flips the state of a target bit conditioned on the state of a control bit. Here we implement a CNOT quantum gate according to the Cirac–Zoller proposal<sup>1</sup>. In our experiment, two <sup>40</sup>Ca<sup>+</sup> ions are held in a linear Paul trap and are individually addressed using focused laser beams<sup>2</sup>; the qubits<sup>3</sup> are represented by superpositions of two long-lived electronic states. Our work relies on

recently developed precise control of atomic phases<sup>4</sup> and the application of composite pulse sequences adapted from nuclear magnetic resonance techniques<sup>5,6</sup>.

Any implementation of a quantum computer (QC) has to fulfil a number of essential criteria (summarized in ref. 7). These criteria include: a scalable physical system with well characterized qubits, the ability to initialize the state of the qubits, long relevant coherence times (much longer than the gate operation time), a qubit-specific measurement capability and a ‘universal’ set of quantum gates. A QC can be built using single qubit operations (‘rotations’) and two-qubit CNOT gates because any computation can be decomposed into a sequence of these basic gate operations<sup>8,9</sup>. So far, quantum algorithms have been implemented only with nuclear magnetic resonance (NMR) and ion trap systems. Recently, using NMR techniques, complex quantum algorithms such as Shor’s factorizing algorithm<sup>10</sup> employing seven qubits have been demonstrated<sup>11</sup>. In NMR systems qubits are encoded in mixed states and ensemble measurements reveal the computational output. Thus, NMR implementations of a QC are not scalable in principle, although they provide an ideal system for testing procedures and algorithms.

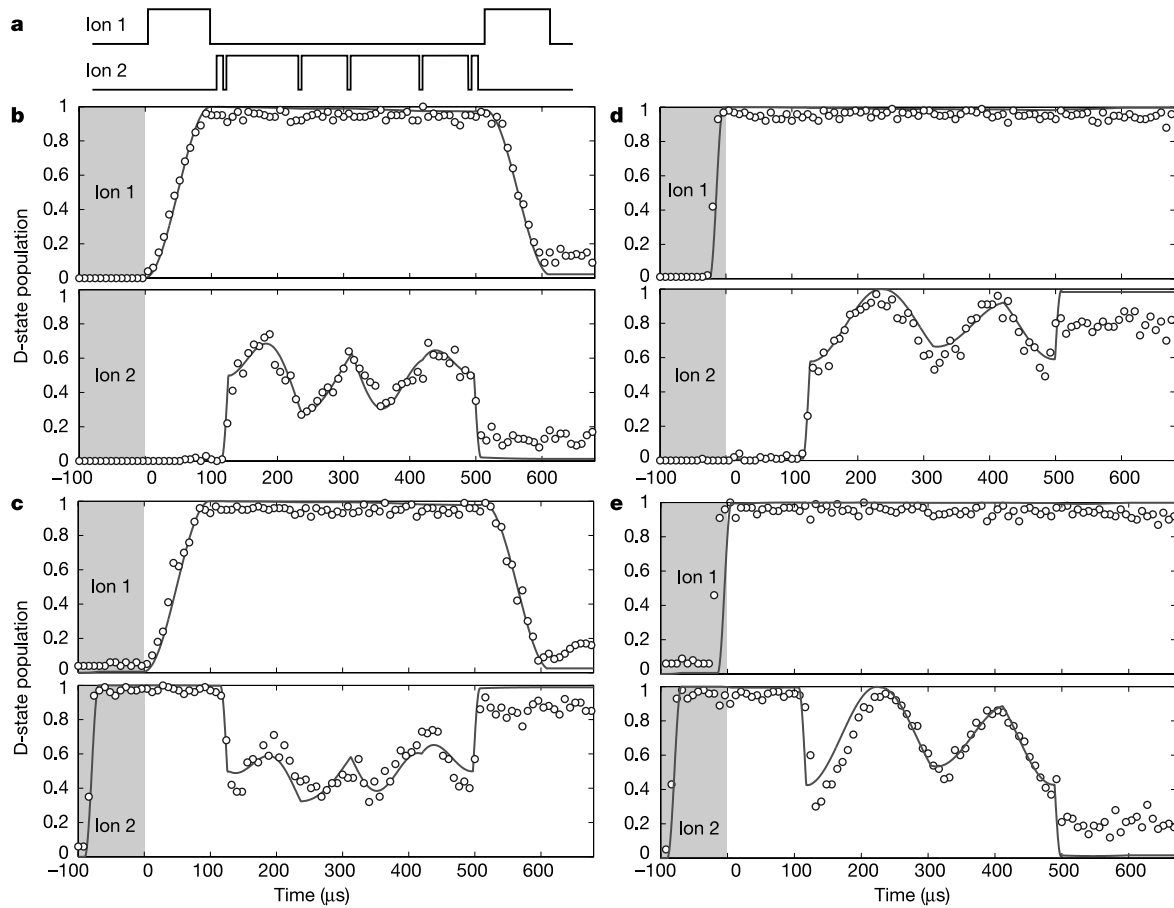
In contrast, trapped ions allow one to prepare and manipulate pure states such that quantum computation is scalable. In addition, cooled, trapped ions exhibit unique features which make them ideally suited for implementations of a QC<sup>12</sup>. In particular, their quantum state of motion can be controlled to the zero point of the trapping potential<sup>13,14</sup>, they provide a long time for manipulations of the qubits<sup>13</sup> encoded in long-lived internal states<sup>3</sup>, and their quantum state can be detected with efficiencies close to 100% (ref. 15). During the past years two and four ions have been entangled<sup>16,17</sup>, a single-ion CNOT gate<sup>18</sup> has been realized, and recently the Deutsch–Jozsa algorithm<sup>19</sup> has been implemented on a single-ion quantum processor. While the entangling operations demonstrated in ref. 17 can be used as a logic gate, a two-ion gate using individual addressing has not been demonstrated. This may serve as a key element for the further development and future perspectives towards general purpose quantum computing with trapped ions.

To implement a QC, Cirac and Zoller proposed a string of ions in a linear trap to serve as a quantum memory where the qubit information is carried by two internal states of each ion. Computational operations are carried out by addressing the ions individually with a laser beam. Single-qubit rotations are performed using coherent excitation by a single laser pulse driving transitions between the qubit states. For a two-qubit CNOT operation, Cirac and Zoller proposed to use the common vibration of an ion string to convey the information for a conditional operation (this vibrational mode is called the ‘bus-mode’). This can be achieved with a sequence of three steps after the ion string has been prepared in the ground state  $\langle n = 0 \rangle$  of the bus-mode. First, the quantum information of the control ion is mapped onto this vibrational mode, that is, the entire string of ions is moving and thus the target ion participates in this common motion. Second, and conditional upon the motional state, the target ion’s qubit is inverted. Finally, the state of the bus-mode is mapped back onto the control ion. Mathematically, this amounts to performing the operation

Table 1 Pulse sequence for the Cirac–Zoller CNOT quantum gate operation

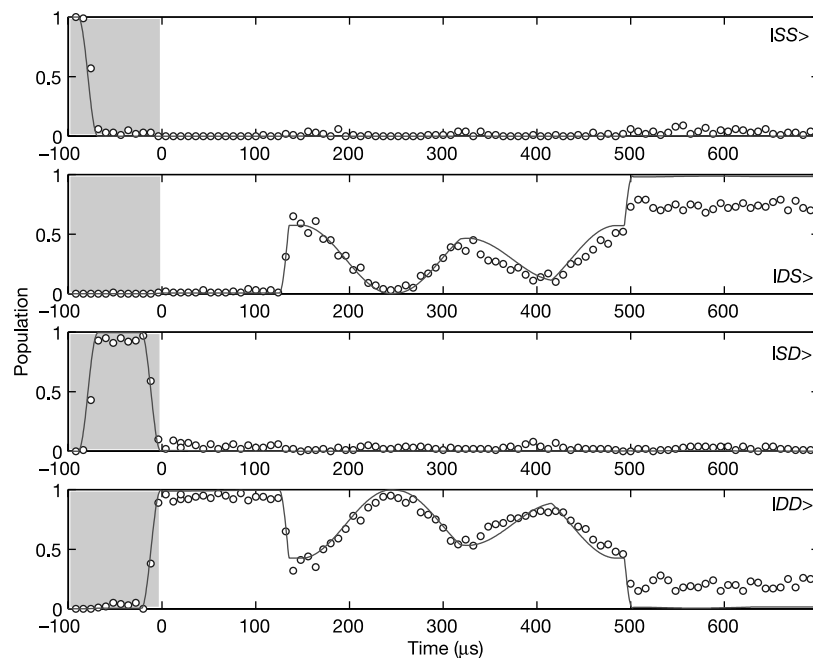
| Ion 1                      |                           | Ion 2           |                                                                 |                 |                              | Ion 1                                     |                                            |
|----------------------------|---------------------------|-----------------|-----------------------------------------------------------------|-----------------|------------------------------|-------------------------------------------|--------------------------------------------|
| $R_1^+(\pi, 0)$<br>Mapping | $R_2(\pi/2, 0)$<br>Ramsey | $R_2^+(\pi, 0)$ | $R_2^+(\pi/\sqrt{2}, \pi/2)$<br>Composite single ion phase gate | $R_2^+(\pi, 0)$ | $R_2^+(\pi/\sqrt{2}, \pi/2)$ | $R_2(\pi/2, \pi)$<br>Ramsey <sup>-1</sup> | $R_1^+(\pi, \pi)$<br>Mapping <sup>-1</sup> |

$R_j(\theta, \varphi)$ ,  $R_j^+(\theta, \varphi)$  denote transitions on the carrier and the blue sideband, respectively, for the  $j$ th ion.  $\theta$  denotes the angle of the rotation, defined by the Rabi frequency and the pulse length;  $\varphi$  denotes the axis of the rotation, given by the phase between the exciting radiation and the atomic polarization.



**Figure 1** State evolution of both qubits under the CNOT operation. For this, the pulse sequence (a) (see also Table 1) is truncated as a function of time and the  $D_{5/2}$  state probability is measured (b–e). The solid lines indicate the theoretically expected behaviour. They do not represent a fit. Input parameters for the calculations are the independently measured Rabi frequencies on the carrier and sideband transitions and the addressing error. The initial state preparation is indicated by the shaded area and drawn in

all figures for negative time values. The actual Cirac–Zoller CNOT gate pulse sequence starts at  $t = 0$ . After mapping the first ion’s state (control qubit) with a  $\pi$ -pulse of length  $95 \mu\text{s}$  to the bus-mode, the single-ion CNOT sequence (consisting of six concatenated pulses) is applied to the second ion (target qubit) for a total time of  $380 \mu\text{s}$ . Finally, the control qubit is reset to its original value with the concluding  $\pi$ -pulse applied to the first ion for  $95 \mu\text{s}$ .



**Figure 2** Joint probabilities for the ions prepared in  $|DD\rangle$ . The data points represent the probability for the ion string to be in the state indicated on the right-hand side. Shaded

areas represent the preparation period. The measurement procedure is the same as in Fig. 1.

Table 2 **Error budget**

| Error source                            | Magnitude                                                                         | Contribution |
|-----------------------------------------|-----------------------------------------------------------------------------------|--------------|
| Laser frequency noise (phase coherence) | ~100 Hz (FWHM)                                                                    | 10%          |
| Laser intensity fluctuations            | ~3% peak to peak                                                                  | ~1%          |
| Laser detuning error                    | ~200 Hz                                                                           | ~2%          |
| Residual thermal excitation             | $\langle n \rangle_{\text{bus}} < 0.02$<br>$\langle n \rangle_{\text{other}} < 6$ | 2%<br>0.4%   |
| Addressing error                        | 5% in Rabi frequency (at neighbouring ion)                                        | 3%           |
| Off-resonant excitations                | for $t_{\text{gate}} = 600 \mu\text{s}$                                           | 4%           |
| Total contribution of error sources     |                                                                                   | ~20%         |

The errors accounted for in our experimental apparatus are specified. Laser frequency and intensity noise occurs as a result of imperfect laser stabilization. Slow drift of the laser locking cavity gives rise to detuning errors. Residual thermal excitation of the bus-mode results from non-optimal sideband cooling. The finite width of the focused laser beam at the position of the ion string produces residual excitation at the site of the non-addressed ion. The intense laser light applied on the blue sideband transitions produces off-resonant carrier excitations. The contributions to the loss of fidelity are calculated from the magnitude of errors occurring from each source. As can be seen from the table, the imperfect locking of our laser is responsible for the majority of the error budget. The individual errors are considered to be independent, the total error results from calculating the success probability given by  $\prod_i (1 - \epsilon_i)$ , where the  $\epsilon_i$  are given by the individual errors below.

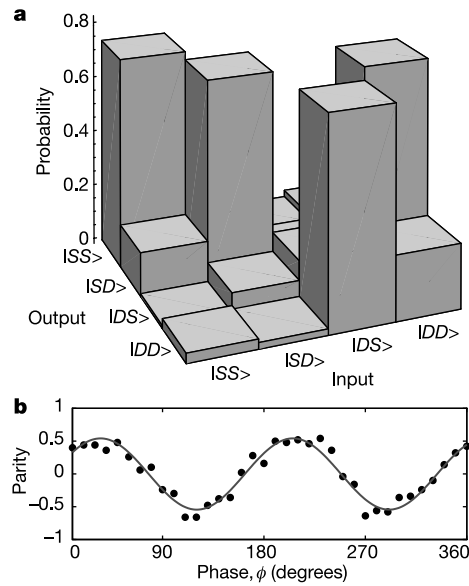
$|\epsilon_1\rangle|\epsilon_2\rangle \rightarrow |\epsilon_1\rangle|\epsilon_1 \oplus \epsilon_2\rangle$  with  $\epsilon_{1,2} \in \{0, 1\}$  describing the logical state of the two qubits in question and  $\oplus$  denoting addition modulo 2. Thus, the CNOT gate operation is described by the unitary operation:

$$\begin{matrix} & |00\rangle & |01\rangle & |10\rangle & |11\rangle \\ \begin{matrix} |00\rangle \\ |01\rangle \\ |10\rangle \\ |11\rangle \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{matrix} \quad (1)$$

where the input and output states of the two qubits  $|\epsilon_1\rangle|\epsilon_2\rangle = |\epsilon_1\epsilon_2\rangle$  are encoded by the electronic states of the ions.

For the experimental implementation we load two  $^{40}\text{Ca}^+$  ions into a linear Paul trap. We encode a qubit in a superposition of the  $S_{1/2}$  ground state and metastable  $D_{5/2}$  state (lifetime  $\tau \approx 1\text{ s}$ )<sup>19</sup> according to  $S_{1/2} \rightarrow |0\rangle$  and  $D_{5/2} \rightarrow |1\rangle$ . The qubits are manipulated on the  $S_{1/2}$  to  $D_{5/2}$  quadrupole transition near 729 nm, using a narrowband Ti:sapphire laser (bandwidth  $< 100\text{ Hz}$ , relative intensity noise  $< 0.02_{\text{r.m.s.}}$ ) which is tightly focused onto individual ions in the string. An electro-optical beam deflector switches the beam between the ions. We measure the qubit by an electron shelving technique<sup>13,15</sup>. For details on the individual state manipulation and detection see the Methods. We start the experimental cycle by Doppler cooling<sup>3</sup> for 2 ms. Using sympathetic sideband cooling<sup>20</sup> for 8 ms we prepare the bus-mode (breathing mode at  $\omega_b = 2\pi \times 2.1\text{ MHz}$ ) in  $|n=0\rangle$ , achieving about 99% ground state occupation. The ions' electronic qubit states are initialized by optical pumping<sup>19</sup>.

Qubit manipulations required for the CNOT operation are realized by applying laser pulses with well-defined phases on the 'carrier' or the 'blue sideband' of the electronic quadrupole transition as described in the Methods and listed in Table 1. Note throughout the following that we always perform sideband pulses on the blue sideband. For the gate sequence, first a  $\pi$ -pulse applied to the blue sideband of the first ion (that is, the control qubit) maps its internal state to a corresponding state of the bus-mode. The phonon-number,  $n$ , of the bus-mode also forms a qubit where  $|n=0\rangle$  ( $|n=1\rangle$ ) represents the logical state  $|1\rangle$  ( $|0\rangle$ )<sup>19</sup>. With the quantum information of the control qubit in the vibrational mode, we address the second ion (that is, the target qubit) and perform a single-ion CNOT gate operation between this ion and the bus-mode. The second ion's internal state is flipped if no vibration is present in the bus-mode, that is, if the bus-qubit is  $|1\rangle$ . This



**Figure 3** Cirac-Zoller CNOT gate operation. **a**, Experimentally observed truth table of the Cirac-Zoller CNOT operation derived from joint-probability measurements as in Fig. 2. Ideally, the table should reproduce the squared moduli of the entries of the unitary operation given in equation (1). Experimentally, we find that the currently available fidelity of the gate operation is limited to about 70–80%; detailed values of measured probabilities are listed below:

|              | $ SS\rangle$ | $ SD\rangle$ | $ DS\rangle$ | $ DD\rangle$ |
|--------------|--------------|--------------|--------------|--------------|
| $ SS\rangle$ | 0.74(3)      | 0.13(3)      | 0.05(3)      | 0.08(3)      |
| $ SD\rangle$ | 0.15(3)      | 0.71(5)      | 0.06(1)      | 0.08(2)      |
| $ DS\rangle$ | 0.01(2)      | 0.08(3)      | 0.14(4)      | 0.77(3)      |
| $ DD\rangle$ | 0.03(3)      | 0.02(1)      | 0.72(6)      | 0.22(4)      |

**b**, Cirac-Zoller gate operation with a superposition  $1/\sqrt{2}(|S\rangle + |D\rangle)|S\rangle$  as input results in an entangled output state  $1/\sqrt{2}(|SS\rangle + e^{i\phi}|DD\rangle)$ . We analyse the entanglement by applying  $\pi/2$ -pulses with phase  $\phi$  to both ions after the gate operation and by measuring the parity<sup>17</sup>  $P = P_{SS} + P_{DD} - (P_{SD} + P_{DS})$  as a function of the phase  $\phi$  ( $P_{ij}$  denotes the probability to find the ions in the states  $|ij\rangle$ ,  $i, j = S$  or  $D$ ). The quantum nature of the gate operation is proved by observing oscillations with  $\cos(2\phi)$ , whereas a non-entangled state would yield a variation with  $\cos(\phi)$  only. From the observed visibility of 0.54(3) and the observed populations  $P_{SS} = 0.42(3)$  and  $P_{DD} = 0.45(3)$  prior to the analysing pulses we calculate<sup>17</sup> a fidelity of  $F = 0.71(3)$ .

operation consists of a pair of Ramsey pulses enclosing a composite phase gate<sup>19</sup>; for details see the Methods. Finally, a  $\pi$ -pulse on the blue sideband applied to the first ion restores the controlling qubit and the bus-mode to their original states. The pulse sequence applied to ions 1 and 2 is sketched in Fig. 1a. The composite pulse sequence replaces the  $2\pi$ -rotation on an auxiliary transition as originally proposed by Cirac and Zoller<sup>1</sup>.

The two ions are prepared in their respective eigenstates, that is, in either the  $|\epsilon_i\rangle = |S\rangle \equiv |0\rangle$  or  $|\epsilon_i\rangle = |D\rangle \equiv |1\rangle$  states using single qubit rotations. In order to trace the state of both qubits under the CNOT operation, we truncate the CNOT pulse sequence at a certain time and measure the probability of finding the ions in the  $D_{5/2}$  state. In Fig. 1b–e we display this probability as a function of time for all four initial settings  $|\epsilon_1\rangle|\epsilon_2\rangle = |\epsilon_1\epsilon_2\rangle = |SS\rangle, |SD\rangle, |DS\rangle, |DD\rangle$ . As shown in Fig. 1, the state evolution of ion 2 follows trajectories depending on the initially prepared qubit states of ion 1. The data agree well with the calculated ideal evolution (given by the solid lines in Fig. 1, no fit parameters). The outcome of the gate operation is inferred from the measured state after the final pulse and it proves



that the second ion's internal state is flipped when the first ion was prepared in the  $|D\rangle$  state (see Fig. 1d, e), whereas it remains in its original state if the first ion was prepared in the  $|S\rangle$  state (see Fig. 1b, c).

In a second experiment, we measured independently for each time step the joint probability of finding the ions in either the  $|SS\rangle$ ,  $|SD\rangle$ ,  $|DS\rangle$  or  $|DD\rangle$  state. Such a joint-probability measurement is displayed in Fig. 2 for the case of the ions starting in the  $|DD\rangle$  state. From these measurements we derived the truth table graphically shown in Fig. 3a. The fidelity of transferring the initially prepared eigenstates into the final target states is optimally 80%; the fidelity of creating an entangled state is  $F = 0.71(3)$ , as shown in Fig. 3b.

The performance of this CNOT gate operation is currently limited by a number of technical shortcomings which are listed in Table 2. The observed fidelity is well understood from these limitations. It is to be expected that technical improvements such as reduction of laser frequency noise and improved control of the addressing beam will allow us to further increase the fidelity to more than 90% in the near future. Still higher fidelities will require more complex composite pulse techniques<sup>21</sup> and/or quantum control techniques<sup>22</sup> which take full account of the experimental imperfections, for example the residual thermal excitation and inaccuracies in the pulse shaping. Eventually, and in order to overcome decoherence for longer pulse sequences, the qubits will have to be encoded in hyperfine states of ions less susceptible to environmental influences, such as  $^{43}\text{Ca}^+$  ions.

These results demonstrate the feasibility and the flexibility of ion trap QC technology. This will provide the basis for experiments using operations between few qubits, as are needed, for example, for the preparation of Bell and GHZ states<sup>23</sup> with ions, for implementing error correction protocols<sup>24</sup>, and for applications with quantum repeaters<sup>25</sup>. □

## Methods

### Level scheme and state manipulation

Quantum information is encoded in trapped  $\text{Ca}^+$  ions employing the electronic  $|S_{1/2}, m_J = -1/2\rangle$  and  $|D_{5/2}, m_J = -1/2\rangle$  levels of a narrow quadrupole transition<sup>3</sup> near 729 nm. The ions are prepared in the ground state of the trap's harmonic oscillator potential by laser cooling<sup>13</sup>. Manipulation of the internal qubit state is performed by exciting the ion on the  $|S_{1/2}\rangle - |D_{5/2}\rangle$  resonance ('carrier') transition while the vibrational degrees of freedom are manipulated on the 'blue detuned sideband', that is, on a transition which changes both the electronic and motional degrees of freedom<sup>19</sup>.

In order to switch between carrier ( $R$ ) and sideband ( $R^+$ ) rotations we shift the laser frequency with an acousto-optical modulator<sup>19</sup>. The phase of the light field is controlled via the phase of the radio frequency driving the acousto-optical modulator with an accuracy of 0.06 rad. Additional phase shifts due to light shifts arise because we have to drive sideband transitions (which couple much more weakly than carrier transitions) with high laser intensity. We cancel the unwanted light shifts with an additional off-resonant laser field inducing a light shift of equal strength but opposite sign<sup>4</sup>.

Phase gate and CNOT gate operations which change the electronic state of a single ion conditional upon the state of motion have been implemented previously using composite pulse techniques<sup>19</sup>. For an implementation of the single-ion CNOT gate operation we use a combination of a pair of Ramsey pulses and a phase gate operation. The computational subspace for this operation consists of the states  $|S,0\rangle$ ,  $|D,0\rangle$ ,  $|S,1\rangle$ ,  $|D,1\rangle$ , that is the electronic states of the ion ( $|S\rangle$ ,  $|D\rangle$ ) and the phonon state ( $|n=0\rangle$ ,  $|n=1\rangle$ ). The matrix describing the phase gate operation is given by:

$$\begin{pmatrix} |S, n=1\rangle & |D, n=1\rangle & |S, n=0\rangle & |D, n=0\rangle \\ |S, n=1\rangle & -1 & 0 & 0 & 0 \\ |D, n=1\rangle & 0 & -1 & 0 & 0 \\ |S, n=0\rangle & 0 & 0 & -1 & 0 \\ |D, n=0\rangle & 0 & 0 & 0 & 1 \end{pmatrix} \quad (2)$$

With a  $2\pi$ -rotation on the blue sideband the  $|S,0\rangle$  and  $|D,1\rangle$  state amplitudes acquire a phase factor of  $-1$ . The state  $|D,0\rangle$  is not affected by this operation and its phase does not change. However, transitions starting from the  $|S,1\rangle$  state would yield a rotation of  $2\pi\sqrt{2}$  because the sideband Rabi frequency depends on the phonon number. This shortcoming can be circumvented by using a combination of four pulses on the target ion possessing different rotation angles (that is, lengths of Rabi pulses) about different rotation axes (that is, phases of the exciting radiation), see Table 1. Together with two carrier  $\pi/2$ -pulses

(Ramsey pulses) before and after this sequence, this completes the single-ion CNOT gate operation.

### Individual ion addressing and read-out

For detection of the internal quantum states, we excite the  $S_{1/2}$  to  $P_{1/2}$  dipole transition<sup>3</sup> near 397 nm and monitor the fluorescence with an intensified charge-coupled device (CCD) camera separately for each ion. Fluorescence indicates that the ion was in the  $S_{1/2}$  state; no fluorescence reveals that it was in the  $D_{5/2}$  state. By repeating the experimental cycle 100 times we find the average state populations. Fluorescence is collected with the CCD camera for 23 ms (data of Fig. 1) and 10 ms (all other data). The fluorescence intensity is integrated over an area of approximately  $3\mu\text{m} \times 3\mu\text{m}$  around the ions by binning the corresponding pixels of the enlarged image on the CCD camera. Thus, state detection of each qubit is performed with an efficiency of about 98% with the residual error of about 2% resulting from spurious fluorescence light of the adjacent ion (cross-talk) or in the case of the 23 ms collection time, from spontaneous decay.

Individual state manipulation of the ions is performed with a tightly focused laser beam near 729 nm. The inter-ion distance is  $5.3\mu\text{m}$  and the gaussian beam width (full-width at half-maximum, FWHM, at the focus) is  $2.5\mu\text{m}$ . Directing the beam onto one ion, the remaining intensity in the wing of the beam incident on the neighbouring ion is suppressed by a factor of  $2.5 \times 10^{-3}$ . The addressing beam can be switched from one ion to the other within  $15\mu\text{s}$  using an electro-optical beam deflector.

Received 18 December 2002; accepted 17 February 2003; doi:10.1038/nature01494.

1. Cirac, J. I. & Zoller, P. Quantum computations with cold trapped ions. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
2. Nägerl, H. C. *et al.* Laser addressing of individual ions in a linear ion trap. *Phys. Rev. A* **60**, 145–148 (1999).
3. Nägerl, H. C. *et al.* Investigating a qubit candidate: Spectroscopy on the  $S_{1/2}$  to  $D_{5/2}$  transition of a trapped calcium ion in a linear Paul trap. *Phys. Rev. A* **61**, 023405 (2000).
4. Häffner, H. *et al.* Precision measurement and compensation of optical Stark shifts for an ion-trap quantum processor. *Phys. Rev. Lett.* (in the press); preprint available at (<http://arXiv.org/abs/physics/0212040>) (2002).
5. Childs, A. M. & Chuang, I. M. Universal quantum computation with two-level trapped ions. *Phys. Rev. A* **63**, 012306 (2001).
6. Levitt, M. H. Composite pulses (NMR spectroscopy). *Prog. Nucl. Magn. Reson. Spectrosc.* **18**, 61–122 (1986).
7. DiVincenzo, D. P. The physical implementation of quantum computation. *Fortschr. Phys.* **48**, 771–783 (2000).
8. Sleator, T. & Weinfurter, H. Realizable universal quantum logic gates. *Phys. Rev. Lett.* **74**, 4087–4090 (1995).
9. DiVincenzo, D. P. Two-bit gates are universal for quantum computation. *Phys. Rev. A* **51**, 1015–1022 (1995).
10. Shor, P. W. in *Proc. 35th Annu. Symp. Foundations of Computer Science* (ed. Goldwasser, S.) 124–133 (IEEE Computer Society Press, Los Alamitos, California, 1994).
11. Vandersypen, L. M. K. *et al.* Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance. *Nature* **414**, 883–887 (2001).
12. Šašura, M. & Bužek, V. Cold trapped ions as quantum information processors. *J. Mod. Opt.* **49**, 1593–1647 (2002).
13. Roos, Ch. *et al.* Quantum state engineering on an optical transition and decoherence in a Paul trap. *Phys. Rev. Lett.* **83**, 4713–4716 (1999).
14. Meekhof, D. M., Monroe, C., King, B. E., Itano, W. M. & Wineland, D. J. Generation of nonclassical motional states of a trapped atom. *Phys. Rev. Lett.* **76**, 1796–1799 (1996).
15. Dehmelt, H. Proposed  $10^{14}\Delta\nu > \nu$  laser fluorescence spectroscopy on  $\text{Ti}^+$  mono-ion oscillator. *Bull. Am. Phys. Soc.* **20**, 60 (1975).
16. Turchette, Q. A. *et al.* Deterministic entanglement of two trapped ions. *Phys. Rev. Lett.* **81**, 3631–3634 (1998).
17. Sackett, C. A. *et al.* Experimental entanglement of four particles. *Nature* **404**, 256–259 (2000).
18. Monroe, C., Meekhof, D. M., King, B. E., Itano, W. M. & Wineland, D. J. Demonstration of a fundamental quantum logic gate. *Phys. Rev. Lett.* **75**, 4714–4717 (1995).
19. Gulde, S. *et al.* Implementation of the Deutsch–Jozsa algorithm on an ion-trap quantum computer. *Nature* **421**, 48–50 (2003).
20. Rohde, H. *et al.* Sympathetic ground state cooling and coherent manipulation with two-ion crystals. *J. Opt. B* **3**, S34–S41 (2001).
21. Jones, J. A. Robust Ising gates for practical quantum computation. *Phys. Rev. A* **67**, 012317 (2003).
22. de Vivie-Riedle, R., Rabitz, H. & Kompa, K. (eds) *Chem. Phys.*, **267**, Issues 1–3 (2001). Special issue on coherent control.
23. Pan, J.-W., Bouwmeester, D., Daniell, M., Weinfurter, H. & Zeilinger, A. Experimental test of quantum nonlocality in three-photon Greenberger–Horne–Zeilinger entanglement. *Nature* **403**, 515–519 (2000).
24. Steane, A. M. Efficient fault-tolerant quantum computing. *Nature* **399**, 124–126 (1999).
25. Duan, L. M., Lukin, M., Cirac, J. I. & Zoller, P. Long-distance quantum communication with atomic ensembles and linear optics. *Nature* **414**, 413–418 (2001).

**Acknowledgements** We thank P. Zoller for discussions and we gratefully acknowledge support by the European Commission (QUEST and QUBITS networks), by the Austrian Science Fund (FWF), and by the Institut für Quanteninformation GmbH. H.H. is supported by the Marie-Curie programme of the European Commission.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to R.B. (e-mail: Rainer.Blatt@uibk.ac.at).

# Experimental demonstration of a robust, high-fidelity geometric two ion-qubit phase gate

D. Leibfried<sup>\*†</sup>, B. DeMarco<sup>\*</sup>, V. Meyer<sup>\*</sup>, D. Lucas<sup>\*‡</sup>, M. Barrett<sup>\*</sup>, J. Britton<sup>\*</sup>, W. M. Itano<sup>\*</sup>, B. Jelenković<sup>\*§</sup>, C. Langer<sup>\*</sup>, T. Rosenband<sup>\*</sup> & D. J. Wineland<sup>\*</sup>

<sup>\*</sup> Time and Frequency Division, National Institute of Standards and Technology, 325 Broadway, Boulder, Colorado 80305, USA

<sup>†</sup> Physics Department, Campus Box 390, University of Colorado, Boulder, Colorado 80309, USA

<sup>‡</sup> Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK

<sup>§</sup> Institute of Physics, PO Box 57, 11001 Belgrade, Serbia-Montenegro

Universal logic gates for two quantum bits (qubits) form an essential ingredient of quantum computation. Dynamical gates have been proposed<sup>1,2</sup> in the context of trapped ions; however, geometric phase gates (which change only the phase of the physical qubits) offer potential practical advantages because they have higher intrinsic resistance to certain small errors and might enable faster gate implementation. Here we demonstrate a universal geometric  $\pi$ -phase gate between two beryllium ion-qubits, based on coherent displacements induced by an optical dipole force. The displacements depend on the internal atomic states; the motional state of the ions is unimportant provided that they remain in the regime in which the force can be considered constant over the extent of each ion's wave packet. By combining the gate with single-qubit rotations, we have prepared ions in an entangled Bell state with 97% fidelity—about six times better than in a previous experiment<sup>3</sup> demonstrating a universal gate between two ion-qubits. The particular properties of the gate make it attractive for a multiplexed trap architecture<sup>4,5</sup> that would enable scaling to large numbers of ion-qubits.

A system of trapped ions interacting with laser radiation is a promising candidate for the implementation of scalable quantum information processors<sup>1</sup>. After demonstrations of the basic interactions necessary for quantum information processing<sup>3,6,7</sup>, including the Sørensen–Mølmer gate<sup>2,8</sup> that is universal for two ion-qubits<sup>2</sup> and has been used to experimentally entangle two and four ion-qubits<sup>3</sup>, our emphasis has shifted to finding reliable ways to scale this approach to many ions and to improve the fidelity of all operations. Scalability may be achieved with a multi-trap architecture<sup>4,5,9,10</sup>. Initial steps toward this goal with ion-qubits that are shuttled between different sub-traps were recently demonstrated<sup>11</sup>. In this Letter, we describe the mechanism and experimental implementation of a geometric phase gate between two ion-qubits. The infidelity of entangled states produced by the gate is about six times smaller than the best result with ions reported so far<sup>3</sup>, making it attractive for future multi-qubit quantum information processing.

The quantum state  $|\Psi\rangle$  of a harmonic oscillator with mass  $m$  and frequency  $\omega$  can be coherently displaced in position–momentum  $(z, p)$  phase space by acting on it with a classical force  $F_0 \sin(\omega t - \phi)$ , resonant with the oscillator frequency<sup>12</sup>. If the force acts for time  $\tau$ , a displacement by  $\Delta z$  and  $\Delta p$  in phase space is formally described by the action of the corresponding displacement operator on  $|\Psi\rangle$  (refs 4, 12, 13):

$$D(\alpha) = \exp[\alpha a^\dagger - \alpha^* a] \text{ with } \alpha = 1/(2z_0)[\Delta z + i\Delta p/(m\omega)]$$

$$= -F_0 z_0/(2\hbar) \exp(i\phi)\tau \quad (1)$$

where  $z_0 = [\hbar/(2m\omega)]^{1/2}$  is the spread of the oscillator's ground

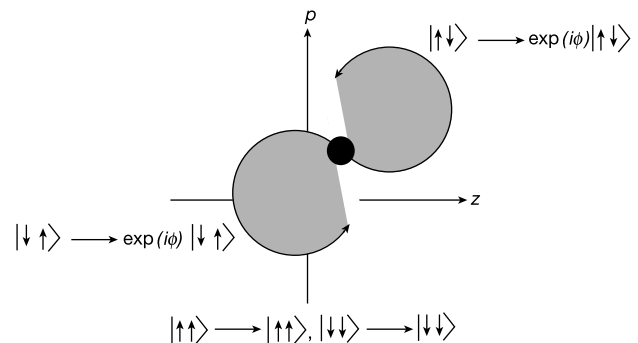
state wavefunction. The effect of two sequential displacements  $D(\alpha)$  and  $D(\beta)$  is additive up to a phase factor:

$$D(\alpha)D(\beta) = D(\alpha + \beta) \exp[i\text{Im}(\alpha\beta^*)] \quad (2)$$

For a suitable sequence of displacements the state  $|\Psi\rangle$  can be transported around a closed loop in  $(z, p)$  phase space. The phase factors for all steps accumulate in such a way that  $|\Psi\rangle$  acquires a geometric phase equal to  $A/\hbar$  if the loop encloses an area  $A$  (see Methods). The acquired phase is independent of the motional state  $|\Psi\rangle$ . As shown below, we can use this geometric phase to realize a logic gate between two ion-qubits if the coherent driving force differs for the two logical states of the ions. In our experiments, we employ a state-dependent dipole force induced by laser light as was used to create 'Schrödinger-cat' states of a single ion<sup>14,15</sup>.

The basic idea for this phase gate was first proposed by Milburn and is a specific case of the more general formalism described by Milburn *et al.*<sup>16</sup>, Sørensen and Mølmer<sup>8</sup> and Wang *et al.*<sup>17</sup> For our particular experiment, the qubit logical states are two  $^9\text{Be}^+$  hyperfine ground states ( $|F=2, m_F=-2\rangle \equiv |\downarrow\rangle$ ,  $|F=1, m_F=-1\rangle \equiv |\uparrow\rangle$ ). We consider a pair of ion-qubits confined together in a harmonic trap potential. The motion of the particles is strongly coupled by their mutual Coulomb repulsion and can be described in terms of normal modes. Along the direction in which the two ions are aligned in the trap (the trap axis) there are two normal modes: the centre-of-mass mode at frequency  $\omega_c$ , where the displacements of both ions from equilibrium are the same, and the 'stretch' mode at frequency  $\omega_s = 3^{1/2}\omega_c$ , where the displacements are equal but in opposite directions.

The state-dependent displacement force is implemented with two laser beams whose frequencies ( $\omega_1, \omega_2$ ) are detuned by  $\Delta \approx +(2\pi) \times 82 \text{ GHz}$  from the  $2s \ ^2S_{1/2} \rightarrow 2p \ ^2P_{1/2}$  electric-dipole transition ( $\lambda \approx 313 \text{ nm}$ ) and have a relative detuning  $\Delta\omega = \omega_1 - \omega_2 = \omega_s + \delta$  close to the frequency of the stretch mode ( $|\omega_s| \gg |\delta|$ ). The electric field from the laser beams gives rise to a Stark shift of the energies of each internal state and a corresponding electric dipole force on each ion. We adjust the polarizations and frequencies of the laser beams to make the average differential energy shift between states  $|\downarrow\rangle$  and  $|\uparrow\rangle$  equal to zero, but on timescales  $1/\Delta \ll t \ll 1/\omega_s$ , there exists a different force on each state, modulated at frequency  $\omega_s + \delta$ . The beam directions are at right angles to each other and the direction of their wavevector difference  $\Delta k = k_1 - k_2$  coincides with the trap axis. The trap potential along this axis was adjusted to fulfil the condition  $\Delta kd = 2\pi p$ , where  $d$  is the distance



**Figure 1** Phase space representation of the stretch-mode amplitude of two trapped ions. The displacement drive moves the motional state components associated with the  $|\uparrow\uparrow\rangle$  and  $|\downarrow\downarrow\rangle$  internal states around circular trajectories in phase space as indicated. Both components acquire the same phase because the enclosed area and sense of rotation are equal.

between the two ions and  $p$  is an integer ( $\omega_s = 2\pi \times 6.1$  MHz,  $\Delta k \approx 2^{1/2} \times 2\pi/(313$  nm) and  $p = 9$  in the experiment). In this case, the interference pattern of the two beams has the same phase at the position of both ions. Therefore, when the ions are in the same internal state the dipole force driving each ion will be the same and no differential force arises, so the stretch mode is not excited. On the other hand, if the ions are in different internal states a differential force exists between them, exciting the stretch mode.

Owing to the detuning  $\delta$ , the driving force is asynchronous with the stretch mode frequency, but re-synchronizes after a duration  $T = 2\pi/\delta$ . During this time the state of the motion is displaced along a circular path in phase space (see Fig. 1 and Methods), returning to the original point in phase space after time  $T$  while acquiring a geometric phase  $\phi = A/\hbar$  equal to the enclosed phase space area. By choosing the intensity of the laser beams appropriately we can achieve  $\phi = \pi/2$ . The evolution of the wavefunction for the two ions can then be summarized by:

$$\begin{aligned} |\downarrow\rangle|\downarrow\rangle|\Psi\rangle &\rightarrow |\downarrow\rangle|\downarrow\rangle|\Psi\rangle \\ |\downarrow\rangle|\uparrow\rangle|\Psi\rangle &\rightarrow e^{i\pi/2}|\downarrow\rangle|\uparrow\rangle|\Psi\rangle \\ |\uparrow\rangle|\downarrow\rangle|\Psi\rangle &\rightarrow e^{i\pi/2}|\uparrow\rangle|\downarrow\rangle|\Psi\rangle \\ |\uparrow\rangle|\uparrow\rangle|\Psi\rangle &\rightarrow |\uparrow\rangle|\uparrow\rangle|\Psi\rangle = e^{-i\pi}(e^{i\pi/2}|\uparrow\rangle)(e^{i\pi/2}|\uparrow\rangle)|\Psi\rangle \end{aligned} \quad (3)$$

The gate is equivalent to a universal controlled  $\pi$ -phase gate plus  $\pi/2$  individual qubit phase shifts on the  $|\uparrow\rangle$  states. In a given algorithm, consisting of phase gates and single-bit rotations, additional single-qubit operations are not necessary to correct for these  $\pi/2$  phase shifts, because they can be absorbed into the next single-bit rotations on the qubits in question.

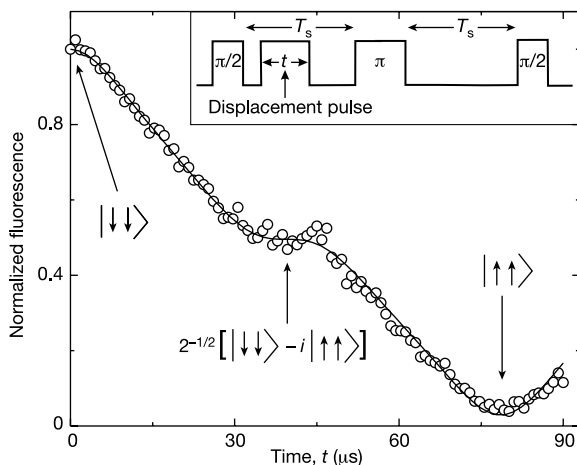
The  $^9\text{Be}^+$  ion-qubits are confined in a linear Paul trap<sup>11</sup>. State preparation is performed by stimulated Raman transitions driven by a pair of laser beams with frequency difference approximately equal to the ground-state hyperfine frequency<sup>6</sup>. When the frequency

difference of the beams is equal to  $\omega_s$ , we drive displacement transitions  $|m_s\rangle|n\rangle \leftrightarrow |m_s\rangle|n+1\rangle$  (that is,  $m_s \in \{\downarrow, \uparrow\}$ ) between oscillator number states  $|n\rangle$  that give rise to the internal-state-dependent displacement. As long as the wavefunction spread and normal mode excursions are smaller than the effective wavelength  $2\pi/\Delta k$  (Lamb–Dicke regime), the displacement only depends on the internal state, and not on the motional state  $|\Psi\rangle$  itself<sup>4</sup>. In our experiment, the ratio of the extension of the stretch-mode ground-state wavefunction for one ion  $z_{0s}$  to the effective wavelength (Lamb–Dicke parameter) was  $\eta_s = \Delta k z_{0s} = 0.19$ . With these conditions and the value of  $\Delta$  noted above we can balance the average a.c. Stark shifts on levels  $|\downarrow\rangle$  and  $|\uparrow\rangle$  and make  $F_{0\downarrow} = -2F_{0\uparrow}$  (ref. 18).

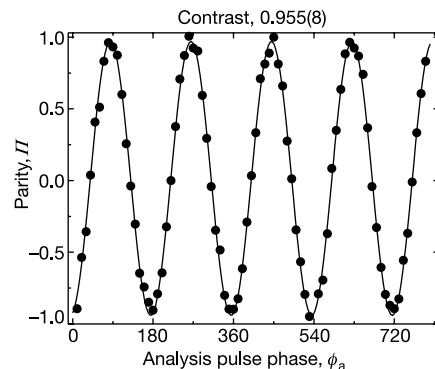
Before each experiment all motional modes were cooled by Doppler laser-cooling, followed by 40 cycles of resolved-sideband Raman cooling that leave the centre-of-mass and stretch mode in the ground state 99% of the time<sup>19</sup>. The two ions were then initialized in the state  $\Psi_0 = |\downarrow\rangle|\downarrow\rangle|n=0\rangle$  by optical pumping. Depending on the experiment, we applied a pulse sequence containing single ion rotations and a state-dependent displacement pulse. We then detected the internal states of both ions through state-dependent resonance fluorescence measurements<sup>20</sup> ( $|\downarrow\rangle$  will strongly fluoresce while  $|\uparrow\rangle$  will produce negligible fluorescence).

It is impossible to observe the effect of the phase gate directly, because what we can observe—the total fluorescence of the two ions—is not affected by the gate operation. Accumulated phases are most conveniently observed in an interferometer experiment. To this end we enclosed the displacement pulse between the first two pulses of a spin-echo pulse sequence acting on both ions: a  $\pi/2$ -pulse, followed by a delay  $T_s$ , a  $\pi$ -pulse, an equal delay  $T_s$  and a final  $\pi/2$ -pulse (see inset in Fig. 2). The spin-echo sequence was used because it makes the qubit transitions immune to the effects of time-varying frequency shifts occurring on timescales much longer than one experimental cycle ( $\geq 300$   $\mu\text{s}$ ), such as those caused by ambient magnetic field fluctuations. Because the displacement pulse and the magnetic field fluctuations are independent (the operators for these two processes commute), the pulse sequence is effectively equivalent to enclosing the displacement pulse in a  $\pi/2$ – $3\pi/2$  Ramsey interference experiment. The first  $\pi/2$ -pulse transforms the state of the ions:

$$\begin{aligned} \Psi_0 = |\downarrow\rangle|\downarrow\rangle|0\rangle &\rightarrow \Psi_1 = 1/2(|\downarrow\rangle|\downarrow\rangle + |\uparrow\rangle|\uparrow\rangle + |\downarrow\rangle|\uparrow\rangle \\ &+ |\uparrow\rangle|\downarrow\rangle)|0\rangle \end{aligned}$$



**Figure 2** State evolution upon displacement. Normalized fluorescence signal (see Methods) after inserting a displacement pulse of variable duration into a spin-echo experiment that is applied to the  $|\downarrow\rangle|\downarrow\rangle|0\rangle$  state (see inset). The motional state returns to its point of origin after every 39  $\mu\text{s}$ , leading to an approximate state  $2^{-1/2}(|\downarrow\rangle|\downarrow\rangle - i|\uparrow\rangle|\uparrow\rangle)|0\rangle$  after 39  $\mu\text{s}$  and to the approximate state  $|\uparrow\rangle|\uparrow\rangle|0\rangle$  after 78  $\mu\text{s}$ . The solid line is a fit to the theoretically expected signal that also allows for an exponential decay in contrast with detuning and decay constant as free parameters. After 39  $\mu\text{s}$ , the fitted decay constant  $\tau_0 = 1.3$  ms predicts a contrast of 0.97, in good agreement with the independently determined fidelity of the entangled state produced for this gate time.



**Figure 3** Parity  $(P_{\downarrow\downarrow} + P_{\uparrow\uparrow}) - (P_{\downarrow\uparrow} + P_{\uparrow\downarrow})$  after producing the maximally entangled state. As  $\phi_a$  is varied, the parity of the two ions should oscillate as  $\cos(2\phi_a)$  with the amplitude of the oscillation equal to twice the magnitude of the density matrix element  $\rho_{\downarrow\downarrow, \uparrow\uparrow}$ . Each data point represents an average of 500 experimental cycles.



With no displacement pulse applied, the subsequent  $\pi$ - and  $\pi/2$ -pulses amount to just a rotation of each ion's internal state by  $3\pi/2$ , so the final state was  $\Psi_F = |\downarrow\rangle|\downarrow\rangle|0\rangle$ , leaving both ions in the fluorescing state. If the displacement pulse is applied for some time  $t < T_S$  after the first  $\pi/2$ -pulse we expect the ions to evolve as:

$$\Psi_1 \rightarrow \Psi_2 = 1/2[ (|\downarrow\rangle|\downarrow\rangle + |\uparrow\rangle|\uparrow\rangle)|0\rangle + e^{i\Phi(t)} (|\downarrow\rangle|\uparrow\rangle)|\alpha(t)\rangle + |\uparrow\rangle|\downarrow\rangle|-\alpha(t)\rangle ]$$

where  $\alpha(t)$  is the amplitude of the motional state displacement on the circular trajectory reached at time  $t$ , and  $\Phi(t)$  is the accumulated phase (see Methods). The subsequent  $\pi$ - and  $\pi/2$ -pulses will in general return the ions not to the original state, but to a more complicated superposition state:

$$\Psi_2 \rightarrow \Psi_3 = 1/2[ (|\downarrow\rangle|\downarrow\rangle + |\uparrow\rangle|\uparrow\rangle)|0\rangle + 1/2 e^{i\Phi(t)} (|\downarrow\rangle|\uparrow\rangle - |\uparrow\rangle|\downarrow\rangle) |\alpha(t)\rangle + 1/2 e^{i\Phi(t)} (|\uparrow\rangle|\downarrow\rangle - |\downarrow\rangle|\uparrow\rangle) |-\alpha(t)\rangle ]$$

fluorescing with a rate:

$$S(t) = S_0 [P_{\downarrow\downarrow} + 1/2(P_{\uparrow\downarrow} + P_{\downarrow\uparrow})] = S_0/2 [1 + \exp(-|\alpha(t)|^2/2) \cos(\Phi(t))] \quad (4)$$

where  $S_0$  is the fluorescence rate of both ions in state  $|\downarrow\rangle$  and  $P_{\downarrow\downarrow}$  is the probability to be in  $|\downarrow\rangle|\downarrow\rangle$ . If  $t = lT$  ( $l$  integer) the motional state returns to its original position in phase space, that is,  $\alpha(lT) = 0$ , and with the intensities of the displacement pulse beams properly adjusted the accumulated phase is  $\Phi(lT) = l\pi/2$ . Figure 2 shows the experimentally observed fluorescence for varying lengths of the displacement pulse (circles). The fitted function (solid line) reflects the theoretical expectation (see Methods) and yields a decrease in contrast of about 3% at the proper gate pulse length of  $T = 39 \mu\text{s}$  (for  $\delta/(2\pi) = 26 \text{ kHz}$ ).

A distinct property of quantum logic gates is that they can produce entanglement out of initially non-entangled states. In particular, inserting a proper  $\pi$ -phase gate pulse between the first two pulses of the spin-echo method would ideally carry out the overall transformation:

$$\Psi_0 = |\downarrow\rangle|\downarrow\rangle|0\rangle \rightarrow \Psi_P = 2^{-1/2} (|\downarrow\rangle|\downarrow\rangle - i|\uparrow\rangle|\uparrow\rangle)|0\rangle \quad (5)$$

thereby creating a maximally entangled state. The entangled nature of the state produced in the experiment can be revealed by applying a final 'analysis'  $\pi/2(\phi_a)$ -pulse, which has a variable phase  $\phi_a$  relative to that of the spin-echo pulses, and deducing the parity:

$$I(\phi_a) = P_{\downarrow\downarrow}(\phi_a) + P_{\uparrow\uparrow}(\phi_a) - [P_{\uparrow\downarrow}(\phi_a) + P_{\downarrow\uparrow}(\phi_a)] \quad (6)$$

from the observed fluorescence histograms<sup>3</sup>. The observed pattern will have a component that oscillates as  $C \cos(2\phi_a)$ , where  $|C|$  is equal to twice the magnitude of the density matrix element  $\rho_{\uparrow\downarrow, \downarrow\uparrow}$  that characterizes the coherence between the  $|\downarrow\rangle|\downarrow\rangle$  and  $|\uparrow\rangle|\uparrow\rangle$  components in the state produced<sup>3</sup>. Figure 3 shows the observed parity signal (circles) plotted versus the phase between the spin-echo pulses and the analysis pulse. The solid line is a cosine fit  $C \cos(b\phi_a)$  to the observed data. The best-fit parameters give frequency  $b = 1.998 \pm 0.002$  and contrast  $|C| = 0.955 \pm 0.008$ . Knowing the contrast of the measured interference pattern, we can directly determine the fidelity  $F = \langle \Psi_P | \rho | \Psi_P \rangle = 1/2(P_{\uparrow\uparrow} + P_{\downarrow\downarrow}) + |\rho_{\uparrow\downarrow, \downarrow\uparrow}|$  of the produced entangled state with density matrix  $\rho$ . Using the populations  $P_{\uparrow\uparrow} + P_{\downarrow\downarrow} = 0.98 \pm 0.02$ , deduced from the histogram decomposition of our entangled-state signal<sup>20</sup> and  $|\rho_{\uparrow\downarrow, \downarrow\uparrow}| = |C|/2 = 0.477 \pm 0.004$ , the measured fidelity is  $F = 0.97 \pm 0.02$ . This is a lower bound for the actual fidelity of the prepared state, because it also includes the infidelity introduced by imperfect

optical pumping to the initial  $|\downarrow\rangle|\downarrow\rangle|0\rangle$  state and imperfect detection.

Although the phase gate is formally the same as the Sørensen–Mølmer gate<sup>8</sup> in a rotated basis, it has some technical advantages. For example, the phase gate has reduced sensitivity to magnetic field fluctuations, because it does not involve spin flips, as was the case for the gates demonstrated previously<sup>3,6</sup>. Because the displacement drive has an interaction strength that is equal to the first-sideband interactions used in the Cirac–Zoller<sup>1</sup> and Sørensen–Mølmer<sup>8</sup> gates (up to factors close to unity), the gate speed for a given laser intensity is about the same. However, for a given fidelity goal, the gate speed can be made greater than in the Cirac–Zoller and Sørensen–Mølmer gates, because the fidelity of those gates is limited by off-resonant contributions of the stronger spin-changing carrier transition<sup>8,21</sup> that is absent in the displacement drive (see Methods).

Like the Sørensen–Mølmer gate, individual ion addressing is not required during the gate (in any of the gates discussed, individual qubit rotations can be accomplished when the ions are spatially separated) and the accumulated phase depends only on the path area, not on the exact starting state distribution, path shape, orientation in phase space, or the time it takes to traverse the closed path. This insensitivity to the starting state means that, within the Lamb–Dicke regime, ground-state cooling is not required for accurate gate operations. The main sources of gate error in our experiment are fluctuations in  $\delta$  and fluctuations in the Raman-beam intensity (both roughly at the 1% level) and a spontaneous emission probability of about 2.2% in each run of the experiment to create the state in equation (5). If frequency drift and intensity errors could be reduced to the order of  $10^{-3}$  and spontaneous emission suppressed (by using a different ion species<sup>18</sup>), the expected gate infidelity is of the order  $10^{-4}$ , the asymptotic threshold value required for fault-tolerant quantum computation<sup>22</sup>. The gate has an additional technical advantage over other gates in the context of quantum computing in multiplexed traps<sup>4,5</sup> where logic ions will be sympathetically cooled with a separate cooling ion before gate operations. In general, the normal-mode amplitudes on each ion will be different, making it technically more difficult to obtain equal laser beam couplings, as required in the Sørensen–Mølmer gate<sup>2,3,8</sup>. Equal coupling is not required for a general geometric phase gate because the extra phases on each qubit can be absorbed into previous or subsequent single-qubit rotations.  $\square$

## Methods

### Geometric phase shift

The geometric phase shift is due to coherent displacement along an arbitrary path. By dividing an arbitrary path into short straight sections  $\Delta\alpha_j, j = \{1, N\}$  and repeatedly invoking equation (2), the total operation can be written as:

$$D_{\text{total}} = D(\Delta\alpha_N) \dots D(\Delta\alpha_1) = D\left(\sum_j \Delta\alpha_j\right) \exp\left[i \text{Im}\left\{\sum_{j=2}^N \Delta\alpha_j \left(\sum_{k=1}^{j-1} \Delta\alpha_k\right)^*\right\}\right] \quad (7)$$

Going to the limit of infinitesimal steps by replacing  $\Delta\alpha_j$  with  $d\alpha$  yields:

$$D_{\text{total}} = D\left[\int d\alpha/d\tau d\tau\right] \exp[i\Phi], \text{ with } \Phi = \text{Im}\left[\int \alpha^*(\tau) [d\alpha/d\tau] d\tau\right] \quad (8)$$

If the path  $P$  is closed, then:

$$D_{\text{total}} = D(0) \exp[i\Phi], \text{ and } \Phi = \text{Im}\left[\int_P \alpha^* d\alpha\right] = 1/\hbar \int_A dz dp = A/\hbar \quad (9)$$

In the last step we converted the path integral over  $P$  to an area integral over the area  $A$  enclosed by  $P$ , and used equation (1) to rewrite  $\alpha$  in terms of position  $z$  and momentum  $p$ .

### Expected signal for the displacement drive

For a relative detuning  $\omega_s + \delta$  of the two Raman beams the infinitesimal displacement of the stretch mode amplitude at time  $\tau$  is<sup>4</sup>:

$$d\alpha_{\uparrow\uparrow}(\tau) = -d\alpha_{\uparrow\downarrow}(\tau) = \Omega_D \exp[i(\phi_L - \delta\tau)] d\tau \quad (10)$$

$$d\alpha_{\downarrow\downarrow}(\tau) = d\alpha_{\downarrow\uparrow}(\tau) = 0$$

where:

$$\Omega_D = -(F_{01} - F_{0\bar{1}})z_{0s}/(2\hbar) \quad (11)$$

$\sigma_0$  is the spread of the ground-state wavefunction for one ion in the stretch mode and  $\phi_L$  is the phase of the driving field. Evaluation of the integrals in equation (8) yields:

$$\alpha_{II}(\tau) = i\Omega_D/\delta[\exp(-i\delta\tau) - 1]\exp(i\phi_L) \quad (12)$$

$$\Phi_{II}(\tau) = (\Omega_D/\delta)^2[\sin(\delta\tau) - \delta\tau]$$

For the  $\pi$ -phase gate the maximum excursion in phase space is  $|\alpha_{II}|_{\max} = 1$ . The normalized fluorescence is equal to:

$$S_{\text{norm}}(\tau) = P_{II}(\tau) + 1/2[P_{II}(\tau) + P_{II}(\tau)] = 1/2[1 + \exp(-|\alpha_{II}(\tau)|^2/2)\cos(\Phi_{II}(\tau))] \quad (13)$$

For the fit in Fig. 3 we allowed for an additional exponential decay:

$$S_{\text{fit}}(\tau) = 1/2[1 + \exp(-\tau/\tau_0)\exp(-|\alpha_{II}(\tau)|^2/2)\cos(\Phi_{II}(\tau))] \quad (14)$$

where  $\tau_0$  is a phenomenological decay constant mimicking decoherence effects.

### Gate speed considerations

The gate speed in our experiments was limited by the intensity of the displacement beams. The observed ratio of gate time  $2\pi/\delta$  to stretch oscillation period  $2\pi\omega_S$  was 238. The largest contribution to gate infidelity from off-resonant excitations is that due to excitation of the centre-of-mass (COM) mode (as opposed to excitation of the internal-state carrier transition in the Cirac–Zoller and Sørensen–Mølmer gates). The ratio of excitation amplitude of the stretch mode to that of the COM mode scales as  $\delta_{\text{COM}}/\delta$  where  $\delta_{\text{COM}} \propto \omega_S$  is the detuning from the COM mode. Therefore if the required fidelity restricts the COM mode excitation to be below a certain size, then the gate rate ( $\propto \delta$ ) will scale linearly with  $\omega_S$ . For our experimental parameters, we estimate the contribution to infidelity from off-resonant excitation of the COM mode to be below  $10^{-4}$ . In a more refined scenario, we could intentionally excite both the COM and the stretch mode. For suitable parameters the COM and stretch-mode amplitudes can return to their initial values even if the gate rate exceeds the trap frequency.

Received 18 December 2002; accepted 7 February 2003; doi:10.1038/nature01492.

1. Cirac, J. I. & Zoller, P. Quantum computations with cold trapped ions. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
2. Sørensen, A. & Mølmer, K. Quantum computation with ions in thermal motion. *Phys. Rev. Lett.* **82**, 1971–1974 (1999).
3. Sackett, C. A. *et al.* Experimental entanglement of four particles. *Nature* **404**, 256–259 (2000).
4. Wineland, D. J. *et al.* Experimental issues in coherent quantum-state manipulation of trapped atomic ions. *J. Res. Natl. Inst. Stand. Technol.* **103**, 259–328 (1998).
5. Kielpinski, D., Monroe, C. & Wineland, D. J. Architecture for a large-scale ion-trap quantum computer. *Nature* **417**, 709–711 (2002).
6. Monroe, C., Meekhof, D. M., King, B. E., Itano, W. M. & Wineland, D. J. Demonstration of a fundamental quantum logic gate. *Phys. Rev. Lett.* **75**, 4714–4717 (1995).
7. Roos, Ch. *et al.* Quantum state engineering on an optical transition and decoherence in a Paul trap. *Phys. Rev. Lett.* **83**, 4713–4716 (1999).
8. Sørensen, A. & Mølmer, K. Entanglement and quantum computation with ions in thermal motion. *Phys. Rev. A* **62**, 02231 (2000).
9. DeVoe, R. G. Elliptical ion traps and trap arrays for quantum computation. *Phys. Rev. A* **58**, 910–914 (1998).
10. Cirac, J. I. & Zoller, P. A scalable quantum computer with ions in an array of microtraps. *Nature* **404**, 579–581 (2000).
11. Rowe, M. A. *et al.* Transport of quantum states and separation of ions in a dual rf ion trap. *Quant. Inf. Comput.* **4**, 257–271 (2002).
12. Carruthers, P. & Nieto, M. M. Coherent states and the forced quantum oscillator. *Am. J. Phys.* **7**, 537–544 (1965).
13. Walls, D. F. & Milburn, G. J. *Quantum Optics* (Springer, Berlin, 1994).
14. Monroe, C., Meekhof, D. M., King, B. E. & Wineland, D. J. A “Schrödinger Cat” superposition state of an atom. *Science* **272**, 1131–1136 (1996).
15. Myatt, C. J. *et al.* Decoherence of quantum superpositions through coupling to engineered reservoirs. *Nature* **403**, 269–273 (2000).
16. Milburn, G. J., Schneider, S. & James, D. F. Ion trap quantum computing with warm ions. *Fortschr. Physik* **48**, 801–810 (2000).
17. Wang, X., Sørensen, A. & Mølmer, K. Multibit gates for quantum computing. *Phys. Rev. Lett.* **86**, 3907–3910 (2001).
18. Wineland, D. J. *et al.* Quantum information processing with trapped ions. Preprint quant-ph/0212079 available at (<http://arXiv.org>) (2002).
19. King, B. E. *et al.* Cooling the collective motion of trapped ions to initialize a quantum register. *Phys. Rev. Lett.* **81**, 1525–1528 (1998).
20. Rowe, M. A. *et al.* Experimental violation of a Bell’s inequality with efficient detection. *Nature* **409**, 791–794 (2001).
21. Steane, A. *et al.* Speed of ion trap quantum information processors. *Phys. Rev. A* **62**, 042305 (2000).
22. Steane, A. Overhead and noise threshold of fault-tolerant quantum error correction. Preprint quant-ph/0207119 available at (<http://arXiv.org>) (2002).

**Acknowledgements** We thank J. Chiaverini, T. Schätz and A. Steane for comments on the manuscript. This work was supported by the US National Security Agency (NSA), the Advanced Research and Development Activity (ARDA). This is a publication of a US government agency.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to D.J.W. (e-mail: david.wineland@boulder.nist.gov).

## An electronic Mach–Zehnder interferometer

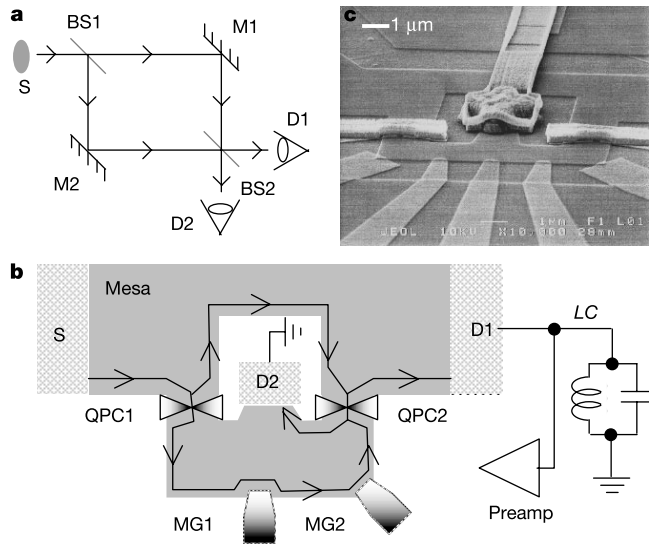
Yang Ji, Yunchul Chung, D. Sprinzak, M. Heiblum, D. Mahalu & Hadas Shtrikman

Braun Center for Submicron Research, Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 76100, Israel

Double-slit electron interferometers fabricated in high mobility two-dimensional electron gases are powerful tools for studying coherent wave-like phenomena in mesoscopic systems<sup>1–6</sup>. However, they suffer from low visibility of the interference patterns due to the many channels present in each slit, and from poor sensitivity to small currents due to their open geometry<sup>3–5,7</sup>. Moreover, these interferometers do not function in high magnetic fields—such as those required to enter the quantum Hall effect regime<sup>8</sup>—as the field destroys the symmetry between left and right slits. Here we report the fabrication and operation of a single-channel, two-path electron interferometer that functions in a high magnetic field. This device is the first electronic analogue of the optical Mach–Zehnder interferometer<sup>9</sup>, and opens the way to measuring interference of quasiparticles with fractional charges. On the basis of measurements of single edge state and closed geometry transport in the quantum Hall effect regime, we find that the interferometer is highly sensitive and exhibits very high visibility (62%). However, the interference pattern decays precipitously with increasing electron temperature or energy. Although the origin of this dephasing is unclear, we show, via shot-noise measurements, that it is not a decoherence process that results from inelastic scattering events.

Direct phase measurements of electrons, customarily done in double-slit interferometers<sup>1–4</sup>, are difficult to perform under strong magnetic fields. Electrons are diverted by the Lorentz force, perform chiral skipping orbits, and prefer one slit to the other—thus breaking the symmetry of the interferometer. At the extreme quantum limit (that is, in the quantum Hall effect, QHE, regime), the skipping orbits quantize to quasi-one-dimensional-like states, named chiral edge states. We have exploited the chiral motion of the electrons, and constructed an electronic analogue of the ubiquitous optical Mach–Zehnder interferometer<sup>9</sup> (Fig. 1a). A beam splitter BS1 splits an incoming monochromatic light beam from source S into two beams, which, after reflection by mirrors M1 and M2, recombine and interfere at BS2 to result in two outgoing beams (collected by detectors D1 and D2). When the phase along one of the paths varies, signals in both D1 and D2 oscillate out of phase, and as no photons are being lost, the sum of both signals stays always equal to the input, S. In the electronic counterpart (Fig. 1b), quantum point contacts (QPCs) function as beam splitters, and ohmic contacts serve as detectors. A QPC is formed in the two-dimensional electron gas (2DEG) by depositing a split metallic gate on the surface of the semiconductor and biasing it negatively with respect to the 2DEG. The induced potential in the 2DEG creates a barrier under the gate bringing the two oppositely propagating edge currents to the small opening in the barrier, thus allowing back-scattering. As shown schematically in Fig. 1b, QPC1 splits the incoming edge current from S to two paths, a transmitted outer path and a reflected inner path; both later recombine and interfere in QPC2, resulting in two edge currents (collected by D1 and D2).

The actual device (Fig. 1c) was fabricated in a high-mobility 2DEG embedded in a GaAs–AlGaAs heterojunction. A ring-shaped mesa, 3  $\mu\text{m}$  in width, was defined by plasma etching with ohmic contacts (for S, D1 and D2) connected to the inner and outer edges of the ring. The inner contact, D2, and the two QPCs are connected to outside sources via metallic films that hover above the surface of

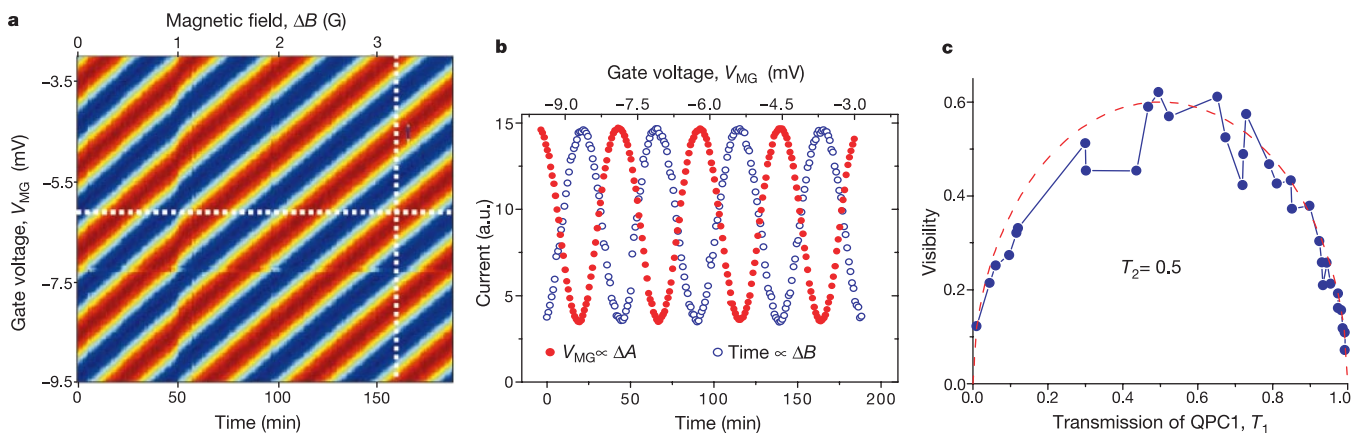


**Figure 1** The configuration and operation of an optical Mach-Zehnder interferometer, and its realization with electrons. **a**, An optical Mach-Zehnder interferometer. D1 and D2 are detectors, BS1 and BS2 are beam splitters, and M1 and M2 are mirrors. With 0 ( $\pi$ ) phase difference between the two paths, D1 measures maximum (zero) signal and D2 zero (maximum) signal. The sum of the signals in both detectors is constant and equal to the input signal. **b**, The electronic Mach-Zehnder interferometer and the measurement system. Edge states are formed in a high, perpendicular, magnetic field. The incoming edge state from S is split by QPC1 (quantum point contact 1) to two paths; one moves along the inner edge, and the other along the outer edge, of the device. The two paths meet again at QPC2, interfere, and result in two complementary currents in D1 and in D2. By changing the contours of the outer edge state and thus the enclosed area between the two paths, the modulation gates (MGs) tune the phase difference between the two paths via the Aharonov-Bohm effect. A high signal-to-noise-ratio measurement of the current in D1 is performed at 1.4 MHz with a cold LC resonant circuit as a band-pass filter followed by a cold, low-noise, preamplifier. **c**, Scanning electron micrograph of the device. A centrally located small ohmic contact ( $3 \times 3 \mu\text{m}^2$ ), serving as D2, is connected to the outside circuit by a long, metallic, air bridge. Two smaller metallic air bridges bring the voltage to the inner gates of QPC1 and QPC2—both serve as beam splitters for edge states. The five metallic gates (at the lower part of the figure) are MGs.

the mesa, called ‘air bridges’. A phase difference  $\varphi$  between the two paths is introduced via the Aharonov-Bohm (AB) effect<sup>10,11</sup>,  $\varphi = 2\pi BA/\phi_0$ , with  $B$  the magnetic field,  $A$  the area enclosed by the two paths ( $\sim 45 \mu\text{m}^2$ ), and  $\phi_0 = 4.14 \times 10^{-15} \text{ T m}^2$  the flux quantum. A few modulation gates, MG, are added above the outer path in order to tune the phase  $\varphi$  by changing the area  $A$ .

We briefly review the operation of the interferometer. At filling factor 1 in the QHE regime, a single chiral edge state carries the current. The interfering current, in turn, is proportional to the transmission probability from source to drain,  $T_{SD}$ . Neglecting dephasing processes and having the transmission (reflection) amplitude  $t_i$  ( $r_i$ ) of the  $i$ th QPC fulfilling  $|r_i|^2 + |t_i|^2 = 1$ , then  $I_{D1} \propto T_{SD1} = |t_1 t_2 + r_1 r_2 e^{i\varphi}|^2 = |t_1 t_2|^2 + |r_1 r_2|^2 + 2|t_1 t_2 r_1 r_2| \cos \varphi$  and  $I_{D2} \propto T_{SD2} = |t_1 r_2 + r_1 t_2 e^{i\varphi}|^2 = |t_1 r_2|^2 + |r_1 t_2|^2 - 2|t_1 t_2 r_1 r_2| \cos \varphi$ , where  $I_{D1}$  and  $I_{D2}$  are the currents in detectors D1 and D2, respectively. Note that ideally the two currents oscillate out of phase as function of  $\varphi$  while  $T_{SD1} + T_{SD2} = 1$ . The visibility of the oscillation is defined as:  $\nu = (I_{\max} - I_{\min})/(I_{\max} + I_{\min})$ , where  $I_{\max}$  and  $I_{\min}$  are the maximum and minimum currents in one of the detectors. For example, when QPC2 is tuned so that  $T_2 = 0.5$ , the visibility is  $\nu = 2\sqrt{T_1(1 - T_1)}$ , where  $|t_i|^2 = T_i$ .

Measurements were done at filling factor 1 (magnetic field  $\sim 5.5 \text{ T}$ ) and also at filling factor 2 with similar results. With a refrigerator temperature of  $\sim 6 \text{ mK}$ , the electron temperature was determined by measuring the equilibrium noise<sup>12</sup> to be  $\sim 20 \text{ mK}$ . High-sensitivity measurements of the interference pattern were conducted at  $\sim 1.4 \text{ MHz}$  with a spectrum analyser. Current at D1 (or D2) was filtered and amplified *in situ* by an LC (inductance + capacitance) circuit and a low-noise, purpose-built pre-amplifier, both placed near the sample and cooled to  $1.5 \text{ K}$ . A standard lock-in technique, with a low-frequency signal ( $7 \text{ Hz}$ ,  $10 \mu\text{V}$  r.m.s.), gave similar results, but the measurement lasted much longer and was affected by the instability of the sample. At  $5.5 \text{ T}$ , each flux quantum occupies an area of some  $10^{-15} \text{ m}^2$  (some 60,000 flux quanta thread the area  $A$ ), so a minute fluctuation in the superconducting magnet’s current or in the area would smear the interference signal. Two measurement methods were used. The first relied on the unavoidable decay of the short-circuited current that circulates in the superconducting magnet, which is in the so-called



**Figure 2** Interference pattern of electrons in a Mach-Zehnder interferometer and the dependence on transmission. **a**, Two-dimensional colour plot of the current collected by D1 as function of magnetic field and gate voltage at an electron temperature of  $\sim 20 \text{ mK}$ . The magnet was set in its persistent current mode ( $B \approx 5.5 \text{ T}$  at filling factor 1 in the bulk) with a decay rate of some  $0.12 \text{ mT h}^{-1}$ , hence time appears on the abscissa. The two QPCs were both set to transmission  $T_1 = T_2 = 0.5$ . Red (blue) stands for high (low) current. **b**, The current (a.u., arbitrary units) collected by D1 plotted as function of the

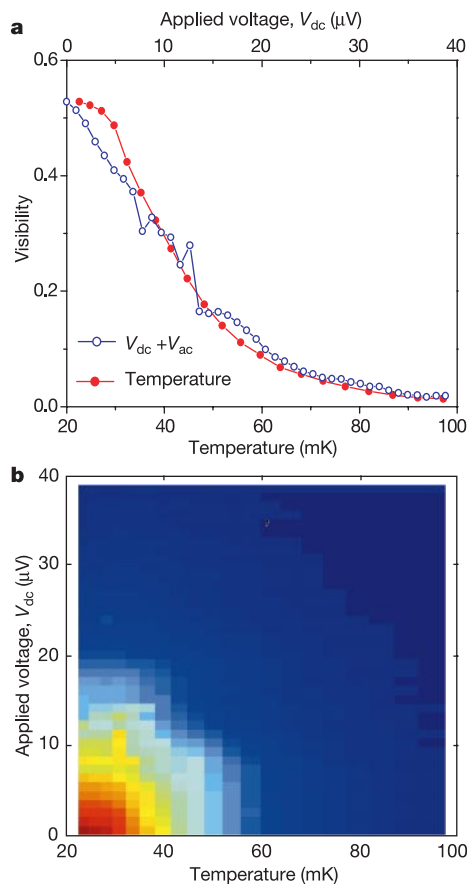
voltage on a modulation gate,  $V_{MG}$  (red plot), and as function of the magnetic field,  $B$  (blue plot)—along the cuts shown in **a**. The visibility of the interference is 0.62. **c**, The visibility of the interference pattern (data points) as a function of the transmission probability  $T_1$  of QPC1 when QPC2 is set to  $T_2 = 0.5$ . Red dashed line is a fit to the experimental data with visibility  $2\eta\sqrt{T_1(1 - T_1)}$ . The normalization coefficient  $\eta = 0.6$  accounts for possible decoherence and/or phase averaging.



persistent current mode. In this mode, the magnetic field decays smoothly at a rate of  $\sim 0.12 \text{ mT h}^{-1}$  ( $\sim 1$  flux quantum every 50 min). The second was via scanning the voltage on a modulation gate at a rate much faster than the decay rate of the magnetic field, thus changing the area  $A$ , the enclosed flux, and consequently the AB phase.

We first test the ideality of the ohmic contacts and the validity of the edge-states picture. For both QPCs open, a nearly ideal Hall plateau was observed in  $I_{D1}$  while no current was measured in D2 ( $I_{D2} = 0$ ). This confirms that current was confined to the outer edge with no backscattering across the  $3\text{-}\mu\text{m}$ -wide mesa. We then pinched off QPC1 or QPC2, and found again a Hall plateau in  $I_{D2}$  with zero current in D1 ( $I_{D1} = 0$ ). This proved that the small ohmic contact of D2 was ideal and fully absorbed the current. Setting then both QPCs to  $T_1 = T_2 \approx 1/2$  and varying the magnetic field  $B$  (actually the time) or the area  $A$  (the voltage on a MG) led to a pronounced interference signal in D1 (or in D2) with visibility as high as 0.62 (Fig. 2). As the field decays linearly with time, and the area (or electron density) varies in proportion to the gate voltage, changing these parameters leads to the diagonal, straight, lines of colour (of constant phase) seen in Fig. 2a. Figure 2b shows similar data taken along two cuts (the dotted lines shown in Fig. 2a)—one for constant  $B$  and one for constant  $A$ . The cleanliness of the interference pattern and the high visibility prove the nearly ideal nature of the interferometer.

In order to verify further the two-path nature of the interference, the visibility was measured as function of  $T_1$  for a constant  $T_2 = 0.5$

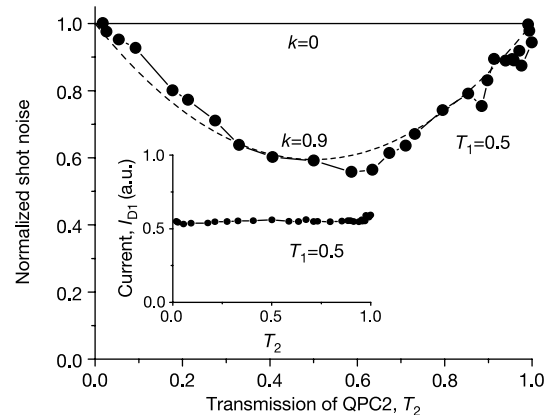


**Figure 3** The dependence of the visibility of the interference pattern on temperature and applied voltage. **a**, Visibility as function of temperature at small excitation voltage for  $V_{dc} = 0$  (red plot), and as function of  $V_{dc}$  with a small a.c. voltage  $V_{ac}$  superimposed on it at electron temperature 20 mK (blue plot). Both QPCs were set to  $T_1 = T_2 = 0.5$ . **b**, A two-dimensional colour plot of the visibility as function of temperature and applied d.c. voltage. Red (blue) stands for high (low) visibility.

(Fig. 2c). It agrees well with the expected expression,  $\nu = 2\eta\sqrt{T_1(1 - T_1)}$ , with  $\eta \approx 0.6$  a normalization factor that accounts for dephasing (either due to phase averaging in the energy window of the electrons, or due to inelastic scattering processes). Moreover, the period of the oscillations, in time and in MG voltage, agrees well with one flux quantum being added (or subtracted) in the ring's area. The time period is  $\sim 50$  min, which is the time needed for one flux quantum decay in the superconducting magnet, while the voltage period agrees approximately with that needed to deplete one electron (hence, one flux quantum for filling factor 1) under the MG gate.

Although the visibility is very high, it is still smaller than unity ( $\nu \approx 0.6$ ). An obvious reason for this is the finite energy spread of the electrons at the edge (due to their finite temperature), and the unavoidable dependence of the AB area on the energy (hence, the AB phase)—leading to phase averaging (thermal smearing). Indeed, the visibility was found to drop precipitously with increasing temperature or applied voltage at S, as seen in Fig. 3. In this example, a mere increase of the temperature to 100 mK (some  $\sim 9 \mu\text{eV}$ ) reduced the visibility from  $\nu \approx 0.53$  to  $\nu \approx 0.01$  (plotted in red in Fig. 3a). If indeed phase averaging is the cause of the dephasing, it could, in principle, be eliminated with monoenergetic electrons. A minute a.c. signal ( $\sim 0.5 \mu\text{V}$ ) at 1.4 MHz was added to a variable d.c. voltage  $V_{dc}$  and the synchronous a.c. part of the interfering signal was measured at 20 mK. This signal leads to a differential visibility  $\nu_d$ , resulting only from the electrons in an energy window  $\sim 0.5 \mu\text{eV}$  around an energy  $eV_{dc}$ . Surprisingly, as seen in Fig. 3a (plotted in blue), the energy-dependent differential visibility at  $T = 20$  mK is similar to the temperature-dependent visibility, with a relation between the scales  $eV_{dc} \approx 4k_B T$ . The visibility (in colour scale) is plotted as function of both  $T$  and  $V_{dc}$  in Fig. 3b. The clear symmetry across the diagonal suggests that the dephasing processes due to temperature and voltage are similar. Unfortunately, this contradicts our previous assertion of phase averaging taking place in a wide window of energy, and points at decoherence, induced by inelastic scattering events, as the main source of dephasing. In other words, for an increased temperature or for high-energy monoenergetic electrons, empty states are being created, allowing energy loss via scattering.

In order to test this hypothesis, current shot noise was measured. Its spectral density is defined as the averaged square of the current



**Figure 4** Shot-noise measurement (at filling factor 2) as function of  $T_2$  when the transmission of QPC1 was set to  $T_1 = 0.5$ . A  $30\text{-}\mu\text{V}$  d.c. voltage (under which the AB interference pattern was quenched) was used to measure shot noise. The shot noise of the current collected by D1 is shown by the black dots to measure shot noise. The shot noise of the current collected by D1 is shown by the black dots (normalized to a maximum), while the two solid lines are the expected noise for  $k = 0$  and  $k = 0.9$ , respectively, according to the simple model described in the text. The agreement with the simple model indicates that the electrons are coherent even at the d.c. voltage where the interference pattern fades away. Inset, the current at D1 as function of  $T_2$  at  $V_{dc} = 30 \mu\text{V}$ . As expected from the lack of the interference pattern, the current is independent of  $T_2$  (see text).

fluctuations per unit of frequency,  $S = \langle i^2 \rangle / \Delta f$ ; for stochastic partitioning at zero temperature,  $S \propto eV_{dc} T_{SD}(1 - T_{SD})$  (ref. 13). Introducing a phenomenological parameter  $k$  that accounts for decoherence in the interferometer with  $T_1 = 1/2$  and  $T_{SD} = 0.5 + k\sqrt{T_2(1 - T_2)}\cos\varphi$ , we find that for complete phase averaging or for a complete decoherence  $T_{SD} = 0.5$ , namely, a constant. On the other hand, shot noise in D1 is  $S_{D1} \propto T_{SD1}(1 - T_{SD1}) = 1/4 - k^2 T_2(1 - T_2)\cos^2\varphi$ , with  $S_{D1} = \text{const.}$  for  $k = 0$ , but  $S_{D1} = 1/4 - k^2 T_2(1 - T_2)/2$  for complete phase averaging (resulting from an integration of  $\cos^2\varphi$  in the range  $\varphi = 0 \dots 2\pi$ ). Hence, noise is expected to exhibit a parabolic dependence on  $T_2$  in a coherent system. Shot noise was measured<sup>12,13</sup> with a relatively large  $V_{dc}$  applied at  $S$  so that interference signal was quenched (negligible visibility). The dependence of  $S$  on  $T_2$ , shown in Fig. 4, followed the above expression with  $k \approx 0.9$ , proving that phase averaging is indeed dominant while decoherence is negligibly small.

A single-particle model (that is, a non-interacting model) would lead to the following dependences of the visibility on energy: for  $V = 0$  and finite  $T$ ,  $\nu \propto \beta T / \sinh(\beta T)$ , with  $\beta$  a constant; for finite  $V$  but  $T = 0$ ,  $\nu \propto \sin[(e/2\pi)V]/[(e/2\pi)V]$ , while the differential visibility at  $T = 0$  is expected to be voltage independent. Because the experimental results contradict these predictions, we propose (with no proof yet) two possible reasons for the dephasing. One might be low-frequency noise (of, say, the  $1/f$  type due to moving impurities), which might be induced by a higher current, leading to fluctuation in the area and consequently, phase smearing. The other could be related to the self-consistent potential contour at the edge. As it depends on the local density of the electrons in the edge state<sup>14</sup>, fluctuation in the density due to partitioning are expected to lead to fluctuation in the AB area enclosed by the two paths and hence to phase randomization. For example, for  $B \approx 5.5$  T, a shift of the edge of only 1–2 Å suffices to add one flux quantum into the enclosed area.

We believe that this electron interferometer might prove useful in future work on the interference of electrons. One possible area of research is the coherence and phase of fractionally charged quasiparticles in the fractional QHE regime<sup>15</sup>. □

Received 1 October 2002; accepted 12 February 2003; doi:10.1038/nature01503.

1. Yacoby, A., Heiblum, M., Umansky, V., Shtrikman, H. & Mahalu, D. Unexpected periodicity in an electronic double slit interference experiment. *Phys. Rev. Lett.* **73**, 3149–3152 (1994).
2. Yacoby, A., Heiblum, M., Mahalu, D. & Shtrikman, H. Coherency and phase sensitive measurements in a quantum dot. *Phys. Rev. Lett.* **74**, 4047–4050 (1994).
3. Schuster, R. *et al.* Phase measurement in a quantum dot via a double-slit interference measurement. *Nature* **385**, 417–420 (1997).
4. Ji, Y., Heiblum, M., Sprinzak, D., Mahalu, D. & Shtrikman, H. Phase evolution in a Kondo-correlated system. *Science* **290**, 779–783 (2000).
5. Buks, E., Schuster, R., Heiblum, M., Mahalu, D. & Shtrikman, H. Dephasing in electron interference by a 'which-path' detector. *Nature* **391**, 871–820 (1998).
6. van der Wiel, W. G. *et al.* The Kondo effect at unitary limit. *Science* **289**, 2105–2108 (2000).
7. Buttiker, M. Four terminal phase coherent conductance. *Phys. Rev. Lett.* **57**, 1761–1764 (1986).
8. Prange, R. E. & Girvin, S. M. (eds) *The Quantum Hall Effect* (Springer, New York, 1987).
9. Born, M. & Wolf, E. *Principles of Optics* 348–352, 7th edn (Cambridge Univ. Press, Cambridge, UK, 1999).
10. Aharonov, Y. & Bohm, D. Significance of electromagnetic potentials in the quantum theory. *Phys. Rev.* **115**, 485–491 (1959).
11. Aronov, A. G. & Sharvin, Yu. V. Magnetic flux effects in disordered conductors. *Rev. Mod. Phys.* **59**, 755–779 (1987).
12. de-Picciotto, R. *et al.* Direct observation of a fractional charge. *Nature* **389**, 162–165 (1997).
13. Reznikov, M. *et al.* Quantum shot noise. *Superlattice Microstruct.* **23**, 901–915 (1998).
14. Chklovskii, D. B., Shklovskii, B. I. & Glazman, L. I. Electrostatics of edge channels. *Phys. Rev. B* **46**, 4026–4034 (1992).
15. Kane, C. L. Telegraph noise and fractional statistics in the quantum Hall effect. Preprint cond-mat/0210621 at (<http://xxx.lanl.gov>) (2002).

**Acknowledgements** We thank Y. Levinson for clarifying the issue of phase averaging, and C. Kane for comments on the manuscript. The work was partly supported by the MINERVA Foundation, the Israeli Academy of Science, the German Israeli Project Cooperation (DIP), the German Israeli Foundation (GIF), and the EU QUACS network.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to M.H. (e-mail: heiblum@wisemail.weizmann.ac.il).

## African vegetation controlled by tropical sea surface temperatures in the mid-Pleistocene period

Enno Schefuß\*, Stefan Schouten, J. H. Fred Jansen & Jaap S. Sinninghe Damsté

Royal Netherlands Institute for Sea Research (NIOZ), PO Box 59, 1790 AB Den Burg, The Netherlands

The dominant forcing factors for past large-scale changes in vegetation are widely debated. Changes in the distribution of  $C_4$  plants—adapted to warm, dry conditions and low atmospheric  $CO_2$  concentrations<sup>1</sup>—have been attributed to marked changes in environmental conditions, but the relative impacts of changes in aridity, temperature<sup>2,3</sup> and  $CO_2$  concentration<sup>4,5</sup> are not well understood. Here, we present a record of African  $C_4$  plant abundance between 1.2 and 0.45 million years ago, derived from compound-specific carbon isotope analyses of wind-transported terrigenous plant waxes. We find that large-scale changes in African vegetation are linked closely to sea surface temperatures in the tropical Atlantic Ocean. We conclude that, in the mid-Pleistocene, changes in atmospheric moisture content—driven by tropical sea surface temperature changes and the strength of the African monsoon—controlled aridity on the African continent, and hence large-scale vegetation changes.

Two main carbon fixation pathways of higher plant photosynthesis, the Calvin–Benson ( $C_3$ ) and the Hatch–Slack ( $C_4$ ) cycles, occur in natural ecosystems<sup>6</sup>. Nearly all trees, cold-season grasses and sedges use the  $C_3$  pathway, whereas  $C_4$  photosynthesis is found in warm-season grasses and sedges<sup>7</sup>. Thus,  $C_4$  plants are found predominantly in tropical savannahs, temperate grasslands and semideserts<sup>7</sup>. Most African grasslands are dominated currently by  $C_4$  plant vegetation<sup>1</sup>.  $C_4$  plants use a  $CO_2$ -concentrating mechanism, thereby outcompeting  $C_3$  plants at low atmospheric  $p_{CO_2}$  (ref. 8) and causing them to be isotopically enriched in  $^{13}C$  (ref. 4). At high  $p_{CO_2}$ , however,  $C_3$  plants will outcompete  $C_4$  plants owing to the higher energy need of  $C_4$  plants during photosynthesis<sup>8</sup>. The crossover between  $C_3$  and  $C_4$  plants depends also on the daytime growing-season temperature, with higher temperatures favouring  $C_4$  plants<sup>7,8</sup>. Other factors, however, may also affect the occurrence of  $C_3$  and  $C_4$  plants. Where precipitation and nutrient availability permit trees to grow,  $C_3$  trees will outcompete  $C_4$  grasses, such as in the tropical forests<sup>1</sup>. The significance of the climatic factors determining the large-scale  $C_4$  plant abundance is still not well understood, but insights can be gained from the analysis of past vegetation changes.

Long-chain, odd-numbered  $C_{25}$  to  $C_{35}$   $n$ -alkanes are major lipid constituents of the epicuticular wax layer of terrestrial plants<sup>9</sup>. These plant waxes are easily removed from the leaf surface by rain or wind, especially by sandblasting during dust storms. They are, therefore, common organic components of eolian dust<sup>10</sup>. In surface sediments of the eastern South Atlantic, the plant wax  $n$ -alkanes exhibit a moderate to high odd versus even carbon-number predominance (carbon preference index (CPI) of 2.3–6.4), and thus predominantly represent leaf waxes of terrestrial higher plants<sup>9</sup>. Their plume-like distribution in surface sediments (Fig. 1a) indicates that they are primarily transported by southeasterly winds from the dry areas in southern Africa, the Kalahari savannah and Namib Desert. The discharge of the Congo River is apparently of minor importance for the supply of  $n$ -alkane leaf waxes. During the austral winter (June to August) the strong Southern Hemisphere trade winds transport

\* Present address: Research Center Ocean Margins, University of Bremen, PO Box 330440, 28334 Bremen, Germany.

large amounts of dust northwestwards over the eastern South Atlantic (Fig. 2a). In the austral summer (December to February), during the southern African monsoon season, the trade winds are deflected landwards (Fig. 2b). We estimated the  $C_4$ -plant-derived percentage of leaf wax  $n$ -alkanes in surface sediments using the stable carbon isotopic composition ( $\delta^{13}C$  in ‰ against Pee Dee Belemnite (PDB)) of the predominating  $n$ - $C_{31}$  alkane and a binary mixing equation, assuming  $\delta^{13}C_4$  leaf wax lipids of  $-21.5\text{‰}$  and  $\delta^{13}C_3$  leaf wax lipids of  $-36.0\text{‰}$  (ref. 11). The results (Fig. 1b) show that the  $\delta^{13}C$  of leaf wax  $n$ -alkanes in the surface sediments of the eastern South Atlantic mainly reflect the contemporary vegetation zones on the adjacent continent (Fig. 2). This finding is supported by investigations of atmospheric dusts collected off the African coast<sup>10</sup>. The atmospheric plant wax lipids show a  $C_4$  plant contribution of about 55% in dust from the southern African dry areas, whereas off the tropical rainforest regions the  $C_4$  plant fraction is 30–40% (ref. 10).

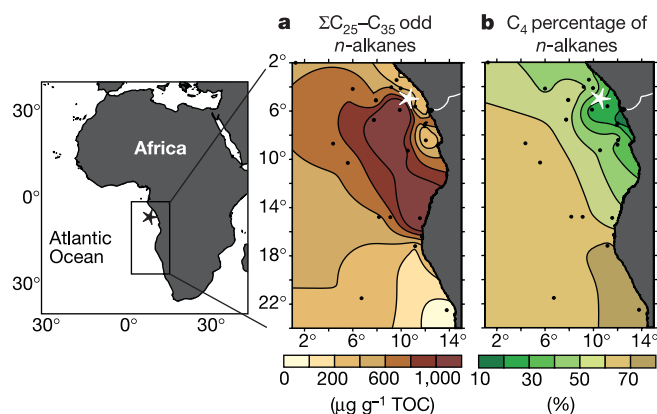
On the basis of these insights into the present-day system, we reconstructed the mid-Pleistocene African  $C_3$  to  $C_4$  plant vegetation changes with a plant wax  $\delta^{13}C$  record from Ocean Drilling Program (ODP) Site 1077 ( $10^\circ 26.2' E$ ,  $5^\circ 10.8' S$ , 2,382 m water depth) in the eastern tropical Atlantic Angola Basin (Fig. 1). This site is located at the northern edge of the modern Southern Hemisphere dust plume during the austral winter (Fig. 2a) and is thus sensitive to changes in its strength and position. It is, therefore, an ideal recorder of the southern African vegetation signal transported with plant wax lipids in dust. This record was compared with an alkenone record obtained from the same sediment core, reflecting the sea surface temperature (SST) (Fig. 3b), and the  $\delta^{18}O$  record of benthic foraminifera at ODP Site 677 (ref. 12), reflecting the global ice volume (Fig. 3d). We investigated the period from 1.2 million years to 450 thousand years before present (kyr BP), including the mid-Pleistocene Transition (MPT), leading to the establishment of the 100-kyr rhythm of the Late Pleistocene ice ages around 650 kyr BP. Before the MPT, a 41-kyr cyclicity was prevalent. During the MPT, from about 920 to 650 kyr BP, the global ice volume variations increased in amplitude<sup>13</sup>, which led to a substantial perturbation of the climate system, probably by re-adjustment of the enlarged Late Pleistocene ice sheets<sup>14</sup>. The global oceanic thermohaline circulation was severely weakened by a strong reduction of North Atlantic deepwater ventilation<sup>15</sup>, probably causing the long-term surface water warming of the tropical Atlantic Ocean during the

MPT, as revealed by our alkenone SST record (Fig. 3b).

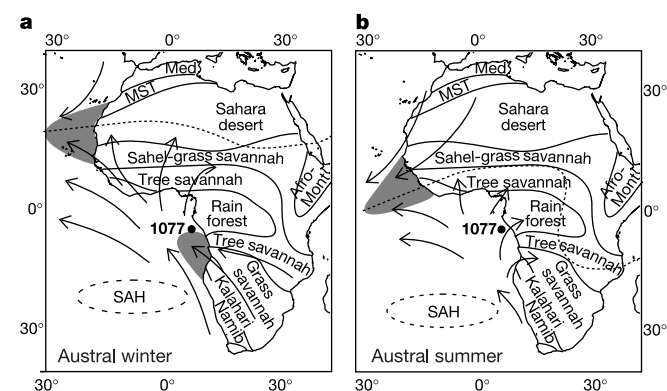
The  $n$ -alkanes in the sediments of ODP Site 1077 are predominantly derived from higher-plant waxes (CPI = 2.1–5.3) supplied by the wind. The mid-Pleistocene record of the relative  $C_4$  plant contribution shows large deviations (from 20% to 70%, Fig. 3a) from the present-day value (about 35%, Fig. 1b). To determine the dominant control on African  $C_4$  plant abundance, cross-spectral analyses of the  $C_4$  plant percentage, SST, global ice volume and  $n$ -alkane flux records with the ETP record—the summed variance of eccentricity, tilt (obliquity) and precession of the Earth's orbit—were performed.

Spectral analyses show that the  $n$ -alkane accumulation rates do not correlate with the  $C_4$  plant changes in all orbital cycles (Table 1). So, an increased eolian flux of  $C_4$  plant  $n$ -alkanes during times of elevated  $n$ -alkane accumulation rates cannot explain the  $C_4$  plant record. The  $n$ -alkane fluxes (Fig. 3c) were generally low before 900 kyr BP, and increased after the growth of the global ice volume at around 900 kyr BP<sup>14</sup>. This suggests that the eolian transport of plant waxes responded to changes in the trade wind system. Conformably, an increase of the eolian terrigenous accumulation off northwest Africa around 900 kyr BP was attributed to the compression and strengthening of the atmospheric circulation cells<sup>16</sup> caused by the expansion of the mean global ice volume. Atmospheric trajectory calculations for this ocean area for the Last Glacial Maximum indicate higher wind speed, but a similar flow path<sup>17</sup>. Therefore, our  $C_4$  plant record reflects predominantly the vegetation changes in southern Africa, and not a change in source area of the wax lipids.

The  $C_4$  plant record correlates strongly with the SST development at ODP Site 1077 (Fig. 3b and Table 1). It co-varies in the precessional (23-kyr) cycle but acts in anti-phase in the obliquity (41-kyr) and eccentricity (100-kyr) cycles. In the precessional cycle, the maximum  $C_4$  plant occurrence corresponds with the maximum SST. This is readily explained by the common low-latitude insolation forcing of the African monsoon and equatorial Atlantic upwelling<sup>16,18,19</sup>, which are both mainly controlled by the strength of the precession<sup>20</sup>. Subtropical southern Africa receives its main precipitation during the austral summer monsoon<sup>21</sup> (Fig. 2b), whereas upwelling in the equatorial Atlantic is strongest during the austral winter<sup>19,22,23</sup> (Fig. 2a). The precessional variations in the summer insolation are in anti-phase between the hemispheres<sup>20,21</sup>, inducing the strongest southern African monsoon when the northern African monsoon is weakest; that is, when equatorial upwelling



**Figure 1** Modern distributions of plant wax concentrations and their vegetation signatures. **a**, **b**, Spatial distribution of the total organic carbon (TOC)-normalized concentrations of the  $C_{25}$  to  $C_{35}$  odd-numbered  $n$ -alkanes (**a**) and the  $C_4$  plant percentage of the  $n$ - $C_{31}$  alkane (**b**) in surface sediments. Isolines were calculated with the Surfer software program using the gridding method of kriging. Black dots are locations of surface samples taken during the RV *Tyro* cruise in 1989. The star indicates the location of ODP Site 1077.



**Figure 2** Present-day atmospheric circulation over Africa, vegetation zones and dust plumes. **a**, **b**, General pattern of present-day atmospheric circulation over Africa and the Southeast Atlantic Ocean in austral winter (**a**; June to August) and austral summer (**b**; December to February). SAH, South Atlantic high-pressure cell. The dotted line indicates the position of the intertropical convergence zone (ITCZ). Dust plumes are shaded. Areas in Africa indicate vegetation zones. The location of ODP site 1077 is indicated. Med, Mediterranean vegetation; MST, Mediterranean–Saharan transitional vegetation; Afro-Mont, Afro-montane vegetation zone.

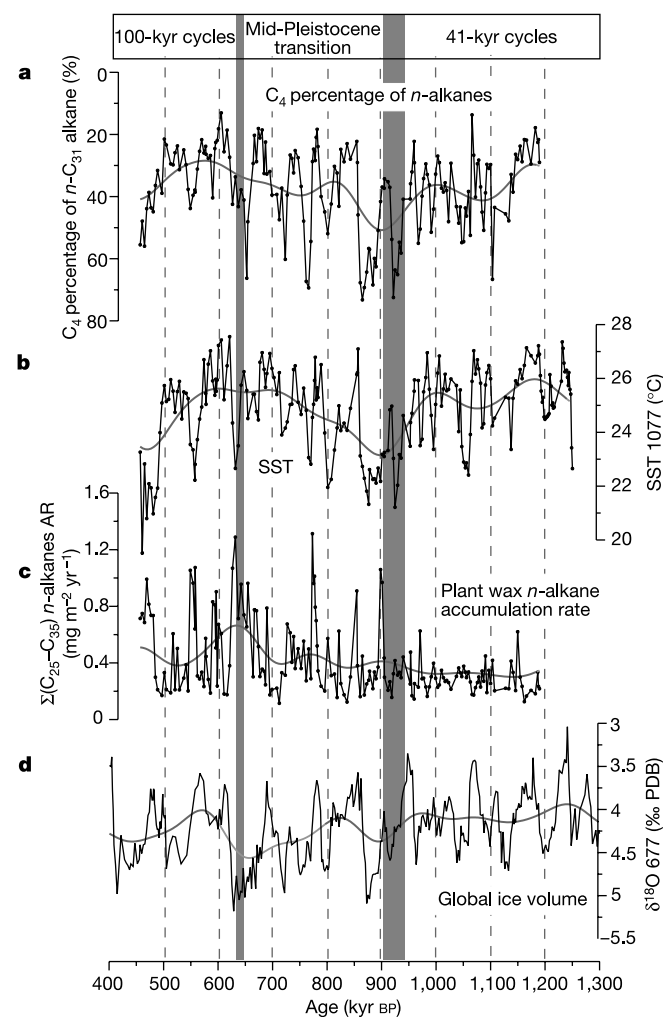


is strongest. The most humid periods in southern Africa, favourable for  $C_3$  plants, therefore coincide with periods of lowered tropical SST<sup>19,21–23</sup> in the precessional cycle. The precessional  $C_3$  to  $C_4$  plant variations in subtropical southern Africa are thus explained by changes in monsoonal precipitation, driven by the low-latitude insolation changes.

The enlarged mean global ice volume, reflected by the  $\delta^{18}O$  of benthic foraminifera (Fig. 3d), increasingly affected the tropical vegetation<sup>24</sup> and the eolian transport (Fig. 3c) after 900 kyr BP. The 100-kyr rhythm of the Late Pleistocene ice ages is detected in all investigated records after about 650 kyr BP. The development of the global ice volume thus undoubtedly influenced the tropical environmental changes, following the frequency pattern set by the high latitudes. In the two longer orbital periods, eccentricity and

obliquity, however, the changes in  $C_4$  plant abundance significantly precede the variations of global ice volume and occur in phase with the tropical SST changes (Table 1). Moreover, the long-term  $C_4$  plant record shows no correlation at all with the long-term global ice volume development (Fig. 3a, d), but it does resemble the long-term SST behaviour (Fig. 3b). The opposite phasing between maximum SST and maximum  $C_4$  plant abundance in the two longer orbital cycles and in the long-term development (Table 1 and Fig. 3) excludes a direct temperature forcing of the observed  $C_3$  to  $C_4$  plant vegetation changes. Assuming that air temperatures over subtropical Africa changed in the same way as the tropical SST, lower temperatures in the growing season would have favoured  $C_3$  plants and not  $C_4$  plants, as is observed (Table 1). We, therefore, propose a common dominant control on large-scale  $C_3$  to  $C_4$  plant variations in southern Africa by continental aridity. In the precessional cycle, the aridity variations were driven by the strength of the monsoon, controlled by low-latitude insolation changes. Low-latitude insolation, however, contains neither an obliquity nor an eccentricity component<sup>20</sup>. This suggests that the tropical SST directly controlled the continental aridity in the longer orbital cycles and the long-term development, in line with recent modelling results<sup>25</sup> and modern climatological observations<sup>26</sup>. During the mid-Pleistocene, the tropical Atlantic SST (Fig. 3b) decreased by up to 5 °C during glacials, comparable to the Last Glacial Maximum<sup>27</sup>. The lowered SST reduced the tropical evaporation and the atmospheric moisture content, the main source for the African precipitation. This large-scale aridification must have been an important factor for the establishment and persistence of  $C_4$  grasslands and even for the evolution of  $C_4$  plants<sup>28</sup>. Enhanced aridity will also have increased the frequency of natural fires, deemed to be an important factor for increasing relative  $C_4$  grass expansion in tropical savannahs<sup>28</sup>. These findings do not necessarily imply an independent behaviour of the  $C_4$  plant abundance from the changes in global ice volume. The high-latitude climate signal manifests itself through its impact on the equatorial Atlantic SST variations<sup>19,27</sup> on the African vegetation changes.

Over the past 420 kyr, the ice-core records of atmospheric  $p_{CO_2}$  correlate strongly with ice volume records<sup>29</sup>. The phase differences between the  $C_4$  plant abundance and the global ice volume record (Table 1) and especially their non-uniform long-term developments indicate, however, that large-scale  $C_3$  to  $C_4$  vegetation changes were not primarily controlled by atmospheric  $p_{CO_2}$  variations. This conclusion is supported by a reconstruction of  $p_{CO_2}$  for the Miocene, when  $C_4$  grasses had widely expanded<sup>7</sup> without an accompanying decrease in atmospheric  $CO_2$  concentrations<sup>2</sup>. Moreover, a low atmospheric  $p_{CO_2}$  alone is unable to trigger  $C_4$  plant expansions, and aridity drives  $C_3$  to  $C_4$  plant abundance changes on a regional scale<sup>3</sup>. Our results indicate that aridity is the dominant climatic control of  $C_4$  plant abundance on a large, continental scale and at various timescales. The variations in Pleistocene African vegetation were primarily forced by changes in the strength of the monsoon and changes in the atmospheric moisture balance directly controlled by the tropical SST. Low-latitude sea-surface conditions



**Figure 3** Mid-Pleistocene changes of African vegetation, tropical Atlantic SST, eolian plant wax transport and global ice volume. Mid-Pleistocene records of: **a**, the  $C_4$  plant fraction of the plant wax  $n$ -alkanes using  $\delta^{13}C$  values of the  $n$ - $C_{31}$  alkane (note axis orientation); **b**, the alkenone-derived SST record at ODP Site 1077; **c**, the accumulation rates (AR) of plant-wax-derived long-chain  $C_{25}$  to  $C_{35}$  odd-numbered  $n$ -alkanes; and **d**, the global ice volume evolution, reflected by the oxygen isotope values of benthic foraminifera at ODP Site 677 (ref. 12) (note axis orientation). The grey lines behind the records are the long-term trends (band-pass filters with a central frequency of 0.004 (250-kyr period) and a bandwidth of 0.004 cycles per kyr). The stratigraphy of ODP Site 1077 is based on oxygen isotope analyses of *Globigerinoides ruber* (pink)<sup>24</sup> by correlation to the stratigraphy of ODP Site 677 (ref. 12). Ages for all samples were determined by linear interpolation using the revised composite depth scale (ref. 30). The division of the age scale into pre-MPT, transition and post-MPT indicated by the grey vertical bars is based on ref. 14.

**Table 1 Results of cross-spectral analyses**

| Data                                   | 1/100 kyr <sup>-1</sup> |      |          |      | 1/41 kyr <sup>-1</sup> |         | 1/23 kyr <sup>-1</sup> |         |
|----------------------------------------|-------------------------|------|----------|------|------------------------|---------|------------------------|---------|
|                                        | $k_0$                   | $k$  | Phi (°)  |      | $k$                    | Phi (°) | $k$                    | Phi (°) |
| $C_4$ percentage                       | 0.96                    | 0.97 | 174 ± 9  | 0.98 | -132 ± 8               | 0.97    | -34 ± 9                |         |
| SST ODP 1077                           | 0.96                    | 0.96 | -8 ± 10  | 0.99 | 48 ± 6                 | 0.94    | -22 ± 13               |         |
| $\Sigma C_{25}-C_{35}$ $n$ -alkanes AR | 0.96                    | 0.85 | 126 ± 21 | 0.94 | 144 ± 14               | 0.99    | 122 ± 4                |         |
| $-\delta^{18}O$ ODP 677                | 0.96                    | 0.98 | 24 ± 8   | 0.99 | 78 ± 5                 | 1.00    | 87 ± 2                 |         |

Variables are crossed with the summed orbital variance record ETP (eccentricity, tilt and precession) based on data from ref. 20. Given are the non-zero coherency at the 80% level ( $k_0$ ), the coherencies ( $k$ ) and phase angles (Phi) with 80% confidence intervals. Positive phase angles indicate that a variable lags the maximum interglacial forcing; negative phases indicate a lead of a variable over maximum interglacial forcing. Benthic  $\delta^{18}O$  data of ODP 677 are from ref. 12. AR, accumulation rate; SST, sea surface temperature.

may thus be far more significant for large-scale environmental changes than previously thought. □

## Methods

Dried and ground sediment samples were ultrasonically extracted with organic solvents. Total lipid fractions (containing the alkenones) were obtained by methylation of the extract, elution over a small silica column using ethyl acetate and silylation. Apolar fractions (containing the *n*-alkanes) were obtained by column chromatography over AgNO<sub>3</sub>-impregnated, activated Al<sub>2</sub>O<sub>3</sub> eluted with hexane:dichloromethane (9:1). Internal standards were added for quantification. Compounds were analysed on a Hewlett Packard 5890 series II gas chromatograph using flame ionization detection (FID) and identified by gas-chromatography mass-spectrometry analyses of selected samples. Quantification of compounds was performed by peak area integration in FID chromatograms. For SST estimation, the simplified unsaturation index ( $U_{37}^{K'}$ ) was calculated from the peak areas of the di- and triunsaturated C<sub>37</sub> alkenones in total lipid fractions. The conversion to SST estimates was done using: SST (°C) = ( $U_{37}^{K'} - 0.044$ )/0.033 (ref. 27). The standard deviation ( $\pm 1$  s.d.) based on duplicate and triplicate analyses of our samples is 0.3 °C. The compound-specific stable carbon isotopic composition of *n*-alkanes was measured through gas chromatography isotope ratio monitoring mass spectrometry using a Finnigan Delta C mass spectrometer. CO<sub>2</sub> gas with known isotopic composition was used as reference. Analyses were done at least in duplicate. Standard deviations of  $\delta^{13}C$  values ( $\pm 1$  s.d.) were better than 0.5‰ against PDB.

Received 28 August 2002; accepted 18 February 2003; doi:10.1038/nature01500.

- Collatz, G. J., Berry, J. A. & Clark, J. S. Effects of climate and atmospheric CO<sub>2</sub> partial pressure on the global distribution of C<sub>4</sub> grasses: present, past and future. *Oecologia* **114**, 441–454 (1998).
- Pagani, M., Freeman, K. H. & Arthur, M. A. Late Miocene atmospheric CO<sub>2</sub> concentrations and the expansion of C<sub>4</sub> grasses. *Science* **285**, 876–879 (1999).
- Huang, Y. *et al.* Climate change as the dominant control on glacial-interglacial variations in C<sub>3</sub> and C<sub>4</sub> plant abundance. *Science* **293**, 1647–1651 (2001).
- Cerling, T. E., Wang, Y. & Quade, J. Expansion of C<sub>4</sub> ecosystems as an indicator of global ecological change in the late Miocene. *Nature* **361**, 344–345 (1993).
- Kuyper, M. M. M., Pancost, R. D. & Sinninghe Damsté, J. S. A large and abrupt fall in atmospheric CO<sub>2</sub> concentration during Cretaceous times. *Nature* **399**, 342–345 (1999).
- O'Leary, M. H. Carbon isotope fractionation in plants. *Phytochemistry* **20**, 553–568 (1981).
- Cerling, T. E. *et al.* Global vegetation change through the Miocene/Pliocene boundary. *Nature* **389**, 153–158 (1997).
- Ehleringer, J. R., Cerling, T. E. & Helliker, B. R. C<sub>4</sub> photosynthesis, atmospheric CO<sub>2</sub>, and climate. *Oecologia* **112**, 285–299 (1997).
- Eglinton, G. & Hamilton, R. J. Leaf epicuticular waxes. *Science* **156**, 1322–1335 (1967).
- Schefe, E., Ratmeyer, V., Stuut, J.-B. W., Jansen, J. H. F. & Sinninghe Damsté, J. S. Carbon isotope analysis of *n*-alkanes in dust from the lower atmosphere over the central eastern Atlantic. *Geochim. Cosmochim. Acta* (in the press).
- Collister, J. W., Rieley, G., Stern, B., Eglinton, G. & Fry, B. Compound-specific  $\delta^{13}C$  analyses of leaf lipids from plants with differing carbon dioxide metabolism. *Org. Geochem.* **21**, 619–627 (1994).
- Shackleton, N. J., Berger, A. & Peltier, W. R. An alternative astronomical calibration of the lower Pleistocene timescale based on ODP Site 677. *Trans. R. Soc. Edinb. Earth Sci.* **81**, 251–261 (1990).
- Shackleton, N. J. & Opdyke, N. O. Oxygen-isotope and paleomagnetic stratigraphy of Pacific core V28–239 late Pliocene to latest Pleistocene. *Geol. Soc. Am.* **145**, 449–464 (1976).
- Mudelsee, M. & Schulz, M. The Mid-Pleistocene climate transition: onset of 100 ka cycle lags ice volume build-up by 280 ka. *Earth Planet. Sci. Lett.* **151**, 117–123 (1997).
- Raymo, M. E., Oppo, D. W. & Curry, W. The mid-Pleistocene climate transition: A deep sea carbon isotopic perspective. *Paleoceanography* **12**, 546–559 (1997).
- deMenocal, P. B. Plio-Pleistocene African climate. *Science* **270**, 53–59 (1995).
- Wypytta, U. & Grieger, B. Comparison of eastern Atlantic atmospheric trajectories for present day and last glacial maximum. *Paleogeogr. Palaeoclimatol. Palaeoecol.* **146**, 53–66 (1999).
- Prell, W. L. & Kutzbach, J. E. Monsoon variability over the past 150,000 years. *J. Geophys. Res.* **92**, 8411–8425 (1987).
- McIntyre, A., Ruddiman, W. F., Karlin, K. & Mix, A. C. Surface water response of the equatorial Atlantic Ocean to orbital forcing. *Paleoceanography* **4**, 19–55 (1989).
- Berger, A. & Loutre, M. F. Insolation values for the climate of the last 10 million years. *Quat. Sci. Rev.* **10**, 297–317 (1991).
- Partridge, T. C., deMenocal, P. B., Lorentz, S. A., Paiker, M. J. & Vogel, J. C. Orbital forcing of climate over South Africa: A 200,000-year rainfall record from the Pretoria saltpan. *Quat. Sci. Rev.* **16**, 1125–1133 (1997).
- Philander, S. G. H. & Pacanowski, R. C. A model of the seasonal cycle in the tropical Atlantic Ocean. *J. Geophys. Res.* **91**, 14192–14206 (1986).
- Katz, E. J. & Garzoli, S. L. Response of the western equatorial Atlantic Ocean to an annual wind cycle. *J. Mar. Res.* **40**, 307–327 (1982).
- Dupont, L. M., Donner, B., Schneider, R. R. & Wefer, G. Mid-Pleistocene environmental change in tropical Africa began as early as 1.05 Ma. *Geology* **29**, 195–198 (2001).
- Ganopolski, A., Rahmstorf, S., Petoukhov, V. & Claussen, M. Simulation of modern and glacial climates with a coupled global model of intermediate complexity. *Nature* **391**, 351–356 (1998).
- Fontaine, B. & Bigot, S. West African rainfall deficits and sea surface temperature. *Int. J. Climatol.* **13**, 271–285 (1993).
- Schneider, R. R., Müller, P. J. & Ruhland, G. Late Quaternary surface circulation in the east equatorial South Atlantic: Evidence from alkenone sea surface temperatures. *Paleoceanography* **10**, 197–219 (1995).
- Sage, R. F. Environmental and evolutionary preconditions for the origin and diversification of the C<sub>4</sub> photosynthetic syndrome. *Plant Biol.* **3**, 202–213 (2001).
- Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
- Jansen, J. H. F. & Dupont, L. M. In *Proceedings of the Ocean Drilling Program, Scientific Results Vol. 175* (ed. Wefer, G. *et al.*) (Ocean Drilling Program, College Station, Texas, 2001).

**Acknowledgements** We thank the Ocean Drilling Program for providing samples. The investigations were supported by the Research Council for Earth and Life Sciences with financial aid from the Netherlands Organisation for Scientific Research.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to E.S. (e-mail: schefuss@uni-bremen.de).

## Fossil evidence for an ancient divergence of lorises and galagos

Erik R. Seiffert\*, Elwyn L. Simons\* & Yousry Attia†

\* Department of Biological Anthropology and Anatomy, Duke University, and Division of Fossil Primates, Duke Primate Center, 1013 Broad Street, Durham, North Carolina 27705, USA

† Egyptian Geological Museum, Misr el Kadima, Ethar el Nabi, Cairo, Egypt

Morphological, molecular, and biogeographic data bearing on early primate evolution suggest that the clade containing extant (or 'crown') strepsirrhine primates (lemurs, lorises and galagos) arose in Afro-Arabia during the early Palaeogene<sup>1</sup>, but over a century of palaeontological exploration on that landmass has failed to uncover any conclusive support for that hypothesis<sup>2</sup>. Here we describe the first demonstrable crown strepsirrhines from the Afro-Arabian Palaeogene—a galagid and a possible lorisid from the late middle Eocene of Egypt, the latter of which provides the earliest fossil evidence for the distinctive strepsirrhine toothcomb. These discoveries approximately double the previous temporal range of undoubted lorisiforms and lend the first strong palaeontological support to the hypothesis of an ancient Afro-Arabian origin for crown Strepsirrhini and an Eocene divergence of extant lorisiform families<sup>1,3</sup>.

The primate clade Strepsirrhini—now represented by the distinctive 'toothcombed' prosimians of the Old World tropics and Madagascar—is one of the three major extant primate groups alongside Anthropoidea (monkeys, apes and humans) and Tarsiiformes (tarsiers). Within Strepsirrhini, it is clear that a major dichotomy exists between a monophyletic Lorisiformes (containing African galagos or 'bushbabies' and African and Asian lorises) and a monophyletic (and wholly Malagasy) Lemuriformes<sup>1,4–6</sup>, but a poor strepsirrhine fossil record has left the age and place of origin of their common ancestor open to debate<sup>7,8</sup>. Given the probable paraphyly of African lorisiforms with respect to Asian lorises<sup>1,3,6</sup>, the proximity of Madagascar to the African mainland<sup>9</sup>, and the distribution of more generalized primates in the Palaeogene fossil record of northern continents and Africa<sup>10</sup>, it is now generally believed that extant strepsirrhines shared a common Afro-Arabian ancestor<sup>1,3</sup>, but the earliest undoubted record of crown Strepsirrhini has long been that of early Miocene (about 20 Myr old) lorisids (lorises) and galagids (galagos) from east Africa<sup>11,12</sup>. These Miocene lorisiforms considerably postdate estimates of basal strepsirrhine divergence times that have been reconstructed using local molecular clocks, which suggest a divergence of lorisiforms and lemuriforms 50–62 Myr ago, and a much wider window of 23–55 Myr for the divergence of lorisids and galagids<sup>1,3,6</sup>.

Palaeontological work in 2001 led to the recovery of two Palaeogene lorisiforms from a single fossil locality, Birket Qarun Locality 2 (BQ-2), that is situated 183 m below the contact of the Qasr el Sagha and Jebel Qatrani formations north of Birket Qarun in the Fayum Depression, Egypt. This horizon was recently placed in the lowermost (Umm Rigi) Member of the Qasr el Sagha Formation<sup>13</sup>, but

renewed work indicates that these fossiliferous alluvial sediments are lithologically distinctive and best retained in the Birket Qarun Formation (T.M. Bown, personal communication), as suggested by earlier research<sup>14,15</sup>. Broadly contemporaneous deposits in the Fayum Depression have been assigned either a basal Priabonian<sup>13</sup> (earliest late Eocene, <36.9 Myr ago) or Bartonian (late middle Eocene, 41.2–36.9 Myr ago)<sup>16</sup> age; we favour the latter interpretation as it is more consistent with palaeomagnetic<sup>17</sup> and faunal<sup>18</sup> evidence from overlying sediments. Specimens described here were recovered either by dry-sieving of loose sediment or by quarrying of bioclastic ironstone conglomerate.

Primates Linnaeus, 1758  
Strepsirrhini Geoffroy, 1812  
Lorisiformes Gregory, 1915  
?Lorisidae Gray, 1821  
*Karanisia* gen. nov.

**Etymology.** In reference to the nearby Ptolemaic/Roman ruins of Karanis.

**Diagnosis.** Differs from all other lorisiforms in exhibiting the following combination of characters: unicuspid lower fourth premolar  $P_4$  with continuous buccal and lingual cingulids, continuous buccal cingulids on lower molars,  $M_3$  longer than  $M_2$  with a large hypoconulid lobe, no  $P^4$  metacone or hypocone, complete buccal and lingual cingula on  $P^4$  through  $M^3$ , upper molars with hypocone lobe, weakly concave distal margins (Fig. 1a) and small crestiform hypocones.

*Karanisia clarki* sp. nov.

**Etymology.** Specific name is in honour of the late anatomist Wilfred Le Gros Clark, who predicted that expeditions to the Fayum region would shed light on early strepsirrhine evolution.

**Holotype.** CGM 40265, left mandible preserving  $M_1$  to  $M_3$  (Fig. 1c).  
**Horizon and locality.** Bartonian Birket Qarun Formation, northern Egypt, locality BQ-2.

**Diagnosis.** As for genus; mean body mass estimate of 273 g (95% confidence interval = 241–304 g) based on  $M_1$  area<sup>19</sup>. For hypodigm, description, and metrics, see Supplementary Information.

Primates Linnaeus, 1758  
Strepsirrhini Geoffroy, 1812  
Lorisiformes Gregory, 1915  
Galagidae Mivart, 1864  
*Saharagalago* gen. nov.

**Etymology.** Generic name is a combination of *sahara*, Arabic word for desert, and the common name for members of the family Galagidae.

**Diagnosis.** Differs from other living and extinct galagids in having having upper molars that are relatively less broad buccolingually, a trenchant prehypocrista coursing from the hypocone to the postprotocrista, and continuous buccal cingulids on lower molars.

*Saharagalago misrensis* sp. nov.

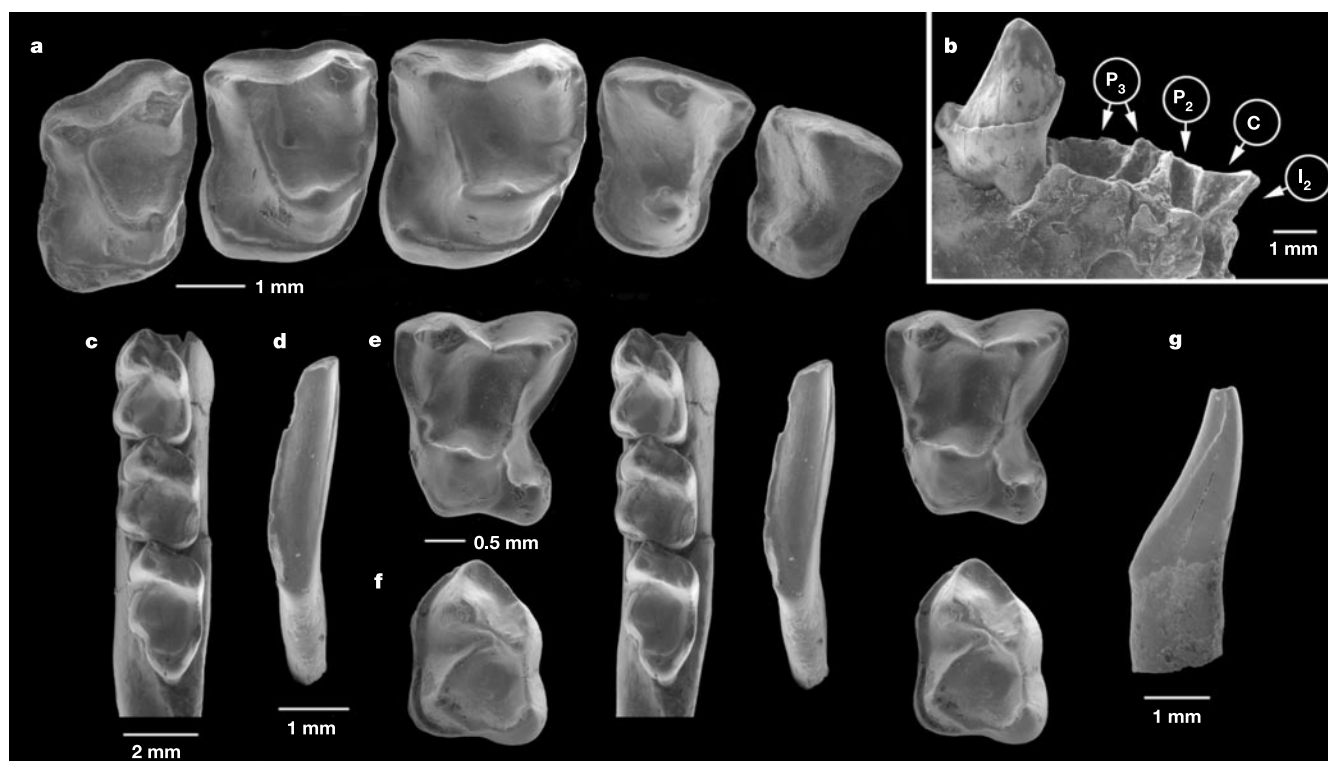
**Etymology.** From *Misir*, Arabic word for Egypt.

**Holotype.** CGM 40266, a left  $M_1$  (Fig. 1f).

**Horizon and locality.** Bartonian Birket Qarun Formation, northern Egypt, locality BQ-2.

**Diagnosis.** As for genus; body mass estimate of 122 g based on  $M_1$  area. For hypodigm, description, and metrics, see Supplementary Information.

An isolated lower canine attributable to *K. clarki* exhibits extreme mesiodistal compression, a grossly elongate and procumbent crown, a mesially inflected root, a lingually restricted distal cingu-



**Figure 1** Middle Eocene lorisiforms from northern Egypt. **a**, isolated upper teeth of *Karanisia clarki* gen et sp. nov.:  $P^3$  (DPC 21371B),  $P^4$  (DPC 21214E, reversed),  $M^1$  (DPC 21639C),  $M^2$  (DPC 21636E),  $M^3$  (DPC 21365C). **b**, right mandible of *K. clarki* (DPC 21249) illustrating size and orientation of damaged alveoli for lower anterior dentition. P, premolar; C, canine; I, incisor. **c–g**, Stereomicrographs of **c**, *K. clarki* holotype (CGM

40265), left mandible preserving  $M_1$  to  $M_3$ ; **d**, isolated left lower canine (DPC 21487B) of *K. clarki*; **e**, *Saharagalago misrensis* gen et sp. nov. probable  $M^1$  (DPC 21214B); **f**, *S. misrensis* holotype  $M_1$  (CGM 40266); **g**, *K. clarki* left lower canine (DPC 21487B), mesial face.

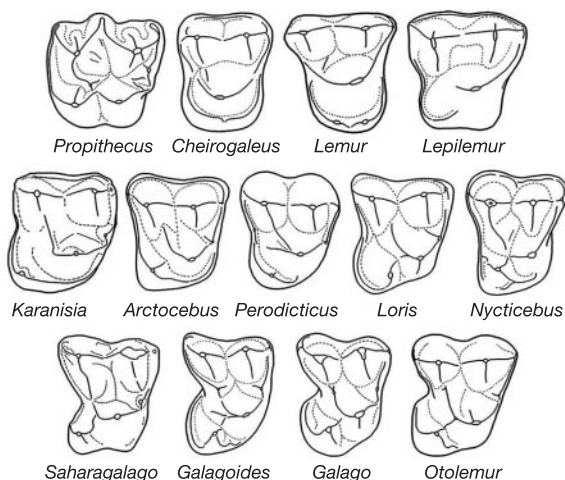


lum, and no distal interstitial wear facet, as is typical of crown strepsirrhine canines (Fig. 1d, g). The size and morphology of the anterior alveoli preserved in a partial mandible of this taxon (Fig. 1b) help to confirm that this tooth was incorporated into a toothcomb like that observable in all extant strepsirrhines aside from the autapomorphic aye-aye (*Daubentonia*). The mesial face of the canine preserves a number of irregularly spaced microscopic grooves coursing perpendicular to the long axis of the crown that are consistent with those created by the passage of hair during grooming<sup>20</sup>. In the upper dentition, *Karanisia* and *Saharagalago* share a constellation of characters (concave distal crown margin delimiting a large hypocone lobe, small paraconule confluent with preprotocrista, no metaconule, trenchant postprotocrista) that is otherwise restricted to Miocene–Recent lorisiforms in the entirety of primate evolutionary history. The upper molar features that *Saharagalago* shares with Miocene–Recent galagids (even greater emargination of the distal crown margin, lingual distension of the hypocone lobe with respect to the lingual face of the protocone, absence of lingual and buccal cingula, long and buccally directed postmetacrista, flaring of the molar margin buccal to the metacone, presence of a prehypocrista) are even more distinctive and of more restricted distribution (Fig. 2).

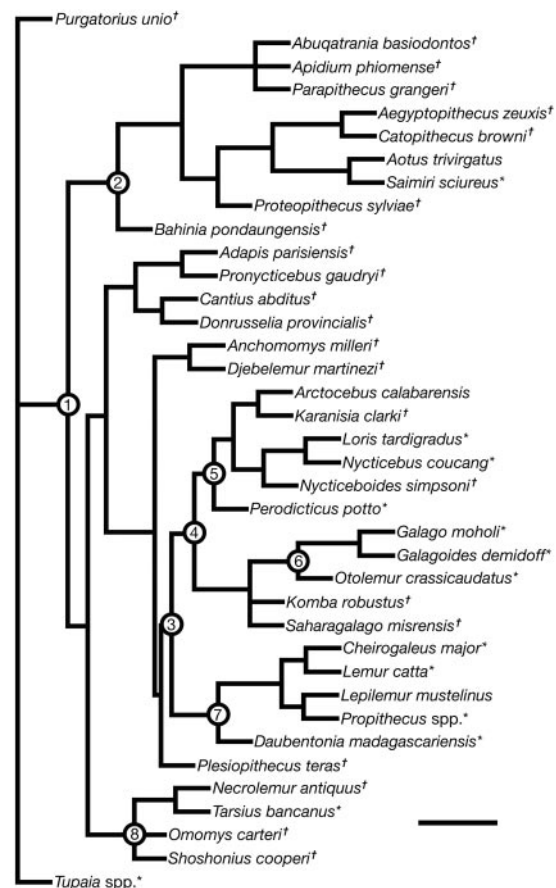
An equally weighted parsimony analysis (Fig. 3) of 2,079 nucleotide characters (previously published<sup>21</sup> mitochondrial cytochrome *b* and nuclear interphotoreceptor retinoid binding protein (IRBP) genes), 279 morphological characters, and one biogeographic character provides reasonably strong support for *Saharagalago*'s galagid status, while *Karanisia* is precariously placed within a poorly supported Lorisidae as the sister taxon of extant *Arctocebus*. The possibility of a *Karanisia*–*Arctocebus* clade might be easily dismissed were it not for the fact that molecular sequence data have uniformly rendered Lorisidae paraphyletic with respect to Galagidae<sup>6,21</sup>, thus implying that divergences among extant lorises (whether paraphyletic or monophyletic<sup>21,22</sup>) are likely to have occurred very shortly after the appearance of crown Lorisiformes. With *Saharagalago* now documenting the divergence of Galagidae from other lorisiforms by the late middle Eocene, we see no compelling basis for rejecting *Karanisia*'s crown lorisid placement on either temporal or available morphological grounds, although missing data and weak bootstrap support certainly leave open the possibility that *Karanisia* is a more basal stem lorisid or stem lorisiform. Unfortunately, this genus sheds little light on the affinities of an isolated primate M<sup>3</sup> from the early Oligocene of Egypt originally identified as lorisiform<sup>23</sup>, but more recently reinterpreted as being anthropoid<sup>24,25</sup> or possibly plesioipithecid<sup>22</sup>; more material of this taxon is needed before any of

these competing hypotheses can be decisively rejected.

The great antiquity of Galagidae implied by these discoveries is consistent with the inferred<sup>1</sup> early Eocene divergence of galagids and lorisids based on the mitochondrial cytochrome *b* gene, but contrasts strongly with the suggested<sup>6</sup> early Miocene divergence of these families based on the nuclear  $\epsilon$ -globin locus. Some interpretations of the scanty lorisiform fossil record have similarly suggested a possible early Miocene divergence of lorisids and galagids<sup>22</sup>, but new fossils from the early Miocene of east Africa appear to attest to a much more ancient split<sup>26</sup>, as do our discoveries. As a means for investigating the possible implications of these ancient fossil lorisiforms for the timing of crown strepsirrhine origins, we employed heuristic quartet dating (using QDate v. 1.11<sup>27</sup>) of the concatenated cytochrome *b* and IRBP data set of ref. 21 using what is presumably a highly conservative<sup>1,3,6</sup> (but palaeontologically unverified) minimum estimate of crown lemuriiform origins equal to our similarly conservative inferred minimum split of galagids and Asian lorisids at 38 Myr ago. The resulting window of minimum estimates for the



**Figure 2** First upper molars of extant strepsirrhines represented in cladistic analysis (modified from Maier<sup>29</sup>) compared with those of *K. clarki* and *S. misrensis* (reversed).



**Figure 3** Strict consensus of 16 equally parsimonious trees recovered in PAUP 4.0 b10 (ref. 30). Data derived from an equally weighted parsimony analysis of a matrix containing 38 taxa and 2,359 molecular, morphological, and biogeographic characters. Branches are scaled relative to bootstrap support (1,000 replicates). Asterisks denote taxa for which cytochrome *b* and IRBP sequences are available; daggers denote fossil taxa. In cases where the majority-rule bootstrap topology resolved polytomies present in the strict consensus topology, those clades are here shown as resolved. Numbered nodes are: (1) crown Primates, (2) Anthropoidea, (3) crown Strepsirrhini, (4) crown Lorisiformes, (5) crown Lorisidae, (6) crown Galagidae, (7) crown Lemuriiformes, and (8) Tarsiiformes. Tree length = 3,807, consistency index excluding uninformative characters = 0.40, retention index = 0.48. Assigning morphological characters weights eight times those of molecular characters produced no change in topology; constraining the tree to recover Haplorhini (a *Tarsius*–Anthropoidea clade excluding Strepsirrhini) did not alter relationships within crown Strepsirrhini. Scale bar represents 100 bootstrap support units.

age of the last common ancestor of lorisiforms and lemuriforms ranged from 49.7 to 53.3 Myr ago (95% confidence interval = 46.9–57.3 Myr ago).

Given that molecular dating has consistently supported an earlier origin for crown lemuriforms than for crown lorisiforms<sup>1,3,6</sup>, and that palaeontological sampling of the early crown strepsirrhine radiation remains woefully inadequate<sup>8,28</sup>, it is remarkable that such early dates are likely to considerably underestimate the true antiquity of the lemuriform–lorisiform split. Nevertheless, by doubling the previous minimum paleontological estimate for the age of crown Lorisiformes, *Karanisia* and *Saharagalago* provide the first convincing fossil evidence attesting to such an ancient origin for crown Strepsirrhini, and further bolster the hypothesis that that clade originated on the Afro-Arabian landmass<sup>1,3</sup>. These taxa also help to inform at least one other major outstanding issue within the clade—the timing of the lemuriform colonization of Madagascar<sup>9</sup>. Although Marivaux *et al.*<sup>7</sup> have recently suggested that early Oligocene *Bugtilemur* from Pakistan is a cheirogaleid lemur specifically aligned with extant *Cheirogaleus* (a conclusion that we must view with skepticism given *Bugtilemur*'s location and age), the current restriction of lemurs to Madagascar would, nevertheless, still be most parsimoniously explicable as the result of a single invasion<sup>1,3</sup>. As molecular data suggest that *Daubentonia* diverged from other crown lemuriforms very shortly after the appearance of crown strepsirrhines<sup>1,3,6</sup>, and before the divergence of galagids from other lorisiforms, it is now reasonable to infer that, at the very latest, lemurs reached Madagascar by the late middle Eocene. □

Received 9 October 2002; accepted 3 February 2003; doi:10.1038/nature01489.

- Yoder, A. D., Cartmill, M., Ruvolo, M., Smith, K. & Vilgalys, R. Ancient single origin for Malagasy primates. *Proc. Natl Acad. Sci. USA* **93**, 5122–5126 (1996).
- Simons, E. L. Discovery of the smallest Fayum Egyptian primates (Anchomomyini, Adapidae). *Proc. Natl Acad. Sci. USA* **94**, 180–184 (1997).
- Yoder, A. D. Back to the future: A synthesis of strepsirrhine systematics. *Evol. Anthropol.* **6**, 11–22 (1997).
- Charles-Dominique, P. & Martin, R. D. Evolution of lorises and lemurs. *Nature* **227**, 257–260 (1970).
- Martin, R. D. *Primate Origins and Evolution: A Phylogenetic Reconstruction* (Princeton Univ. Press, Princeton, 1990).
- Porter, C. A. *et al.* Phylogeny and evolution of selected primates as determined by sequences of the e-globin locus and 5' flanking regions. *Int. J. Primatol.* **18**, 261–295 (1997).
- Marivaux, L. *et al.* A fossil lemur from the Oligocene of Pakistan. *Science* **294**, 587–591 (2001).
- Martin, R. D. Origins, diversity, and relationships of lemurs. *Int. J. Primatol.* **21**, 1021–1049 (2000).
- McCall, R. A. Implications of recent geological investigations of the Mozambique Channel for the mammalian colonization of Madagascar. *Proc. R. Soc. Lond. B* **264**, 663–665 (1997).
- Rose, K. D., Godinot, M. & Bown, T. M. in *Anthropoid Origins* (eds Fleagle, J. G. & Kay, R. F.) 1–28 (Plenum, New York, 1994).
- Simpson, G. G. The tertiary lorisiform primates of Africa. *Bull. Mus. Comp. Zool.* **136**, 39–62 (1967).
- Phillips, E. M. & Walker, A. in *The Primate Fossil Record* (ed. Hartwig, W. C.) 83–95 (Cambridge Univ. Press, Cambridge, 2002).
- Gingerich, P. D. Marine mammals (Cetacea and Sirenia) from the Eocene of Gebel Mokattam and Fayum, Egypt: Stratigraphy, age, and paleoenvironments. *Univ. Mich. Pap. Paleontol.* **30**, 1–84 (1992).
- Beadnell, H. J. L. *The Topography and Geology of the Fayum Province of Egypt* (Survey Department of Egypt, Cairo, 1905).
- Bown, T. M. & Kraus, M. J. Geology and paleoenvironment of the Oligocene Jebel Qatrani Formation and adjacent rocks, Fayum Depression, Egypt. *U S Geol. Surv. Prof. Pap.* **1452**, 1–64 (1988).
- Holroyd, P. A., Simons, E. L., Bown, T. M., Polly, P. D. & Kraus, M. J. New records of terrestrial mammals from the upper Eocene Qasr el Sagha Formation, Fayum Depression, Egypt. *Palaeovertebrata* **25**, 175–192 (1996).
- Kappelman, J., Simons, E. L. & Swisher, C. C. III New age determinations for the Eocene–Oligocene boundary sediments in the Fayum Depression, northern Egypt. *J. Geol.* **100**, 647–668 (1992).
- Simons, E. L., Rasmussen, D. T., Bown, T. M. & Chatrath, P. S. in *Anthropoid Origins* (eds Fleagle, J. G. & Kay, R. F.) 179–201 (Plenum, New York, 1994).
- Conroy, G. C. Problems of body-weight estimation in fossil primates. *Int. J. Primatol.* **8**, 115–137 (1987).
- Rose, K. D., Walker, A. C. & Jacobs, L. L. Function of the mandibular tooth comb in living and extinct mammals. *Nature* **289**, 583–585 (1981).
- Yoder, A. D., Irwin, J. A. & Payseur, B. A. Failure of the ILD to determine data combinability for slow loris phylogeny. *Syst. Biol.* **50**, 408–424 (2001).
- Rasmussen, D. T. & Nekaris, K. A. Evolutionary history of lorisiform primates. *Folia Primatol.* **69** Suppl. 1, 250–285 (1998).
- Simons, E. L., Bown, T. M. & Rasmussen, D. T. Discovery of two additional prosimian primate families (Omomyidae, Lorisidae) in the African Oligocene. *J. Hum. Evol.* **15**, 431–437 (1986).
- Godinot, M. in *Anthropoid Origins* (eds Fleagle, J. G. & Kay, R. F.) 235–295 (Plenum, New York, 1994).
- Gheerbrant, E., Thomas, H., Sen, S. & Al-Sulaimani, Z. Nouveau primate Oligopithecinae (Simiiformes) de l'Oligocène inférieur de Taqah, Sultanat d'Oman. *CR Acad. Sci. Ser. IIa* **321**, 425–432 (1995).
- MacLatchy, L. & Kityo, R. A lower Miocene lorisid femur from Napak, Uganda. *Am. J. Phys. Anthropol.* **34** Suppl., 104–105 (2002).

- Rambaut, A. & Bromham, L. Estimating divergence dates from molecular sequences. *Mol. Biol. Evol.* **15**, 442–448 (1998).
- Tavaré, S., Marshall, C. R., Will, O., Soligo, C. & Martin, R. D. Using the fossil record to estimate the age of the last common ancestor of extant primates. *Nature* **416**, 726–729 (2002).
- Maier, W. Konstruktionsmorphologische Untersuchungen am Gebiß der rezenten Prosimiae (Primates). *Abh. Senckenb. Naturforsch. Ges.* **538**, 1–158 (1980).
- Swofford, D. L. *PAUP\*, Phylogenetic Analysis Using Parsimony and Other Methods* (Sinauer, Sunderland, Massachusetts, 1998).

**Supplementary Information** accompanies the paper on Nature's website (<http://www.nature.com/nature>).

**Acknowledgements** We thank K. Soliman, A. Swedan, A. Hassen, F. Imbabi, and the numerous other scientists from the Egyptian Geological Survey and Mining Authority and the Egyptian Geological Museum who have supported and facilitated our collaborative palaeontological work in northern Egypt. P. Chatrath managed field operations in Egypt. T. M. Bown discovered BQ-2, and together with M. Mathison measured the geological section reported herein. T. M. Bown, P. Chatrath and M. Mathison also helped with collecting during the 2001 season along with A. Abd el-Ghafar, M. Abd el-Ghany, Y. Abd el-Razik, H. Ali, J. Fleagle, G. Kamel, A. Hassen, M. Hassen Taha, C. Heesy, L. Jolly, P. Lewis, G. Mohamed, S. Mohamed, R. Patnaik and M. Zakaria. We thank F. Ankel-Simons, R. Martin, J. Rossie, T. Ryan, C. Simons and A. Yoder for discussion and/or comments on the manuscript. Recent work in Egypt has been generously supported by a National Science Foundation grant to E.L.S. and Leakey Foundation grants to E.R.S.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to E.R.S. (e-mail: erik.seiffert@duke.edu).

## Earliest known crown-group salamanders

Ke-Qin Gao\*† & Neil H. Shubin†

\* School of Earth and Space Sciences, Peking University, Beijing 100871, China

† Department of Organismal Biology and Anatomy, University of Chicago, 1027 E. 57th Street, Chicago, Illinois 60637, USA

Salamanders are a model system for studying the rates and patterns of the evolution of new anatomical structures<sup>1–4</sup>. Recent discoveries of abundant Late Jurassic and Early Cretaceous salamanders are helping to address these issues<sup>5–8</sup>. Here we report the discovery of well-preserved Middle Jurassic salamanders from China, which constitutes the earliest known record of crown-group urodeles (living salamanders and their closest relatives). The new specimens are from the volcanic deposits of the Jiulongshan Formation (Bathonian)<sup>9–13</sup>, Inner Mongolia, China, and represent basal members of the Cryptobranchidae, a family that includes the endangered Asian giant salamander (*Andrias*) and the North American hellbender (*Cryptobranchus*). These fossils document a Mesozoic record of the Cryptobranchidae, predating the previous record of the group by some 100 million years<sup>14–17</sup>. This discovery provides evidence to support the hypothesis that the divergence of the Cryptobranchidae from the Hynobiidae had taken place in Asia before the Middle Jurassic period.

Amphibia Linnaeus, 1758

Lissamphibia Haeckel, 1866

Caudata Scopoli, 1777

Urodela Dumeril, 1806

Cryptobranchioidea Dunn, 1922

Cryptobranchidae Fitzinger, 1826

*Chunerpeton tianyiensis* gen. et sp. nov.

**Holotype.** Chinese Academy of Geological Sciences (CAGS)-IG-02051, natural mould of dorsal and ventral aspects of an articulated



skeleton including the cranium and postcranium (Figs 1 and 2).

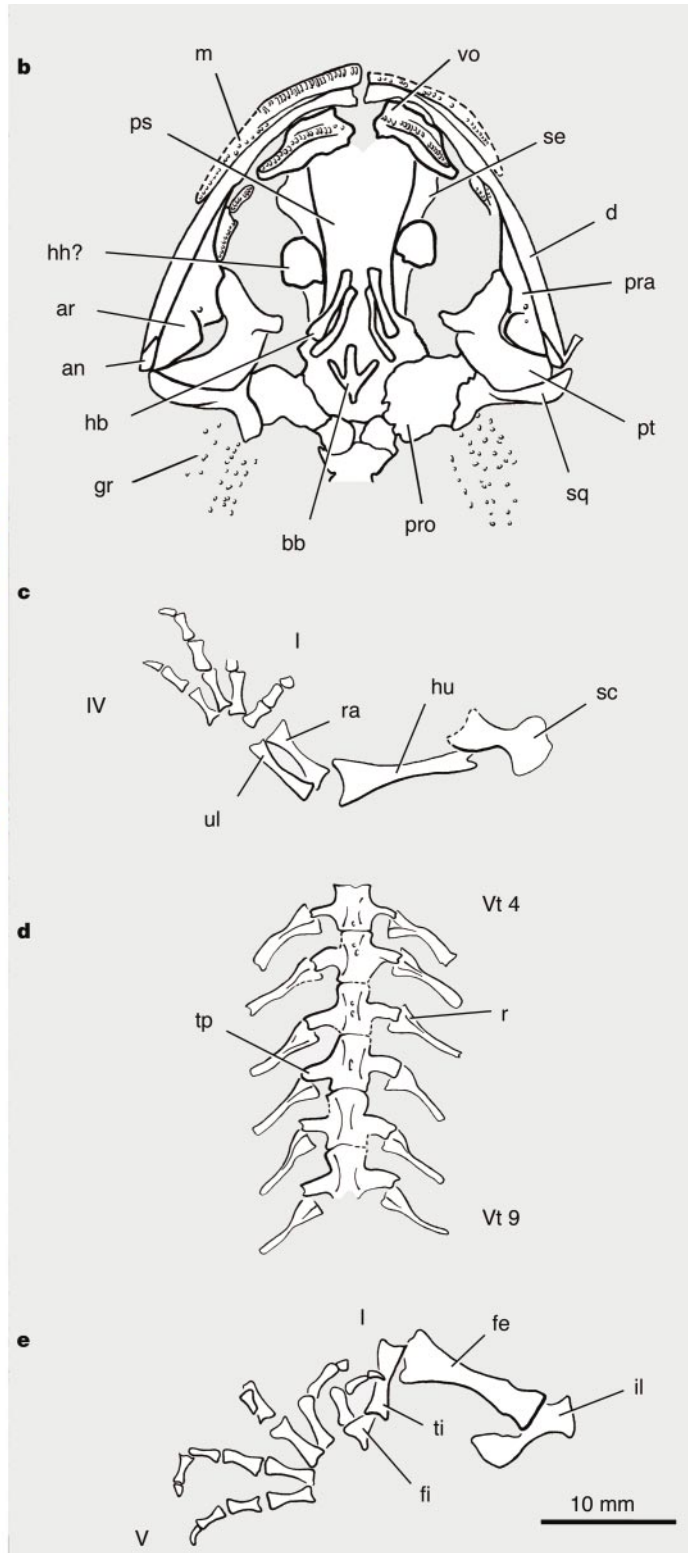
**Etymology.** Chu (Pinyin) meaning early and herpeton (Greek) meaning creeping animal; tianyi, ancient country name for Ningcheng.

**Locality and horizon.** Daohugou, Ningcheng County, Inner Mongolia, China; Middle Jurassic Jiulongshan Formation (Bathonian, 161 million years (Myr) ago)<sup>9–13</sup>.

**Referred material.** Peking University Palaeontology Collections

(PKUP) V0210-0212, specimens preserved as natural moulds of articulated skeletons.

**Diagnosis.** *Chunerpeton tianyiensis* shares with living cryptobranchoids derived characters including: presacral vertebrae bearing unicapitate ribs; reduction in the number of rib-bearing anterior caudal vertebrae reduced to two or three. *Chunerpeton tianyiensis* shares with cryptobranchids derived characters such as: nasal much



**Figure 1** Holotype (CAGS-IG-02051) of *Chunerpeton tianyiensis* gen. et sp. nov. **a**, Natural mould of ventral aspect of entire specimen. **b**, Line drawing of skull in palatal

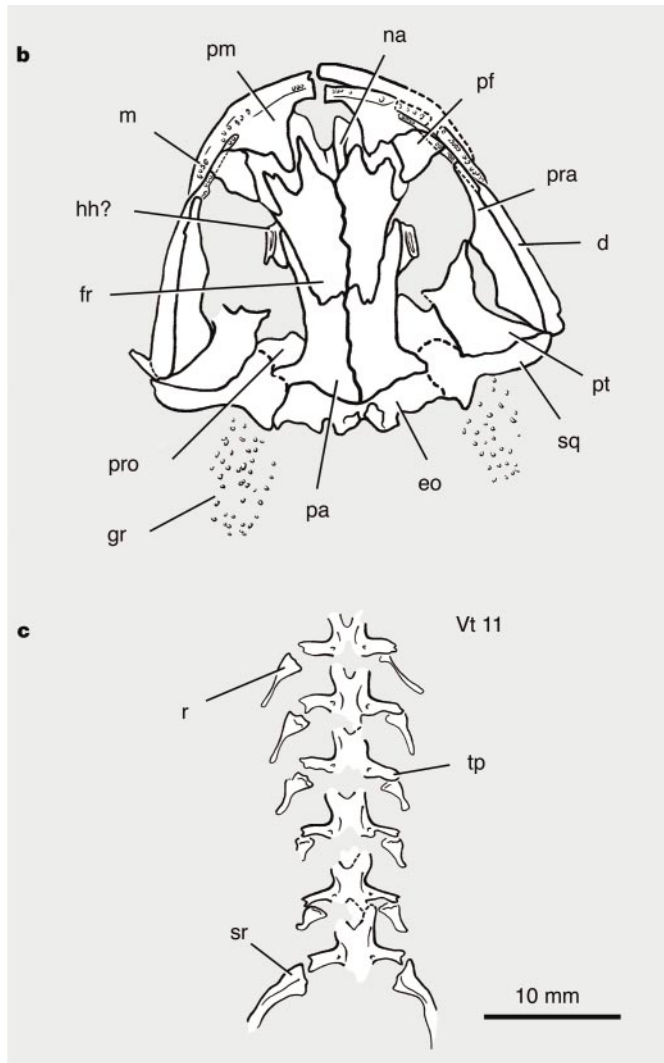
view. **c**, Line drawing of left pectoral girdle and forelimb. **d**, Ventral aspect of anterior trunk vertebrae. **e**, Details of left pelvis and hindlimb. See Fig. 2 for abbreviations.



narrower than interorbital width; nasal–prefrontal contact absent; frontals extend anteriorly to lateral border of nasal; lachrymal absent; anterolateral process of parietal extends along lateral border of frontal; internal carotid foramina penetrate palatal surface of parasphenoid. *Chunerpeton tianyiensis* differs from extant cryptobranchids in lacking midline contact of dorsal processes of premaxillae; frontal–maxillary contact absent; absence of contact between anterolateral process of parietal and prefrontal; vomers without posterior extension; retention of palatal fenestra between

vomers; presence of distinct medial process of pterygoid; pterygoid–parasphenoid contact absent; basibranchial II ossified and trident-shaped; first three pairs of ribs with spatulate distal end; phalangeal formula of 2-2-3-(3/4)-3 in pes.

The new salamander fossils were recovered from volcanic deposits of pale-grey shales and tuffs at Daohugou Village, Inner Mongolia, China. Virtually all of the 200 specimens recovered from the site are larval or subadult individuals with articulated skeletons, many of which preserve evidence of soft tissues. The assessment of



**Figure 2** Holotype (CAGS-IG-02051) of *C. tianyiensis* gen. et sp. nov. **a**, Natural mould of dorsal aspect of entire specimen. **b**, Line drawing of skull in dorsal view. **c**, Line drawing of posterior trunk vertebrae in dorsal view. an, angular; ar, articular; bb, basibranchial; d, dentary; eo, exoccipital; fe, femur; fi, fibula; fr, frontal; gr, gill rakers; hb, hypobranchial;

hh?, hypohyal?; hu, humerus; il, ilium; m, maxilla; na, nasal; pa, parietal; pf, prefrontal; pm, premaxilla; pra, prearticular; pro, pro-otic; ps, parasphenoid; pt, pterygoid; r, rib; ra, radius; sc, scapula; se, sphenethmoid; sq, squamosal; sr, sacral rib; ti, tibia; tp, transverse process; ul, ulna; vo, vomer; vt, vertebra.



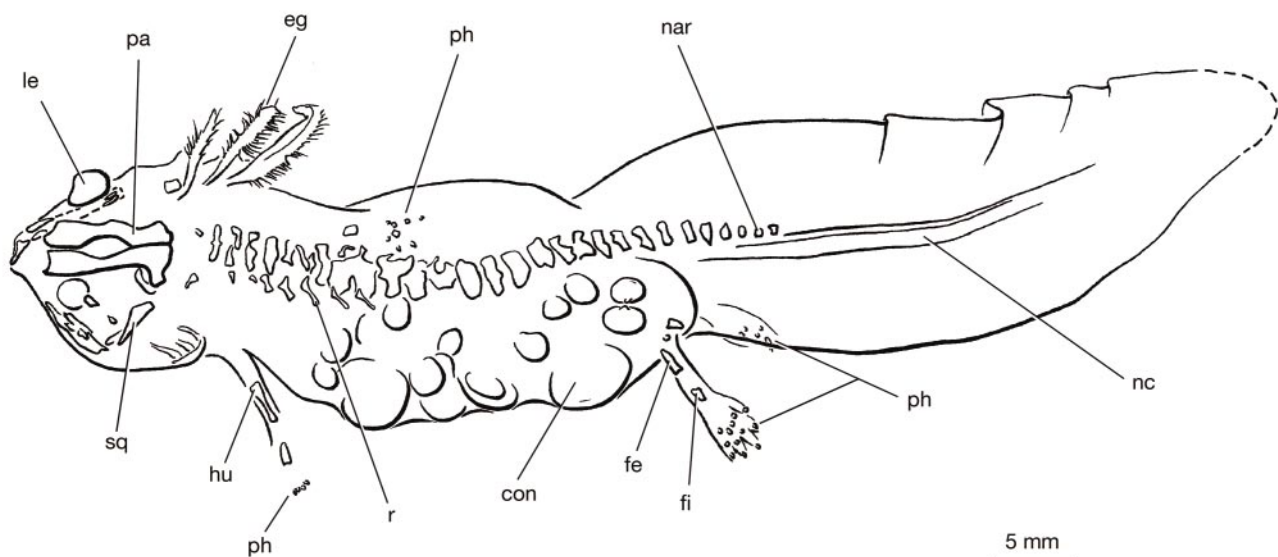
the Middle Jurassic (Bathonian) age of the fossil beds is based on biostratigraphic analysis of insect and vertebrate assemblages<sup>9–13</sup>. Therefore, these fossils represent the earliest known relatives of living salamanders, as other Middle Jurassic forms (such as *Kokartus* and *Marmorerpeton*) lie outside the crown group<sup>18,19</sup>.

The holotype of the new taxon has a total length of approximately 180 mm. Despite its large size, this specimen retains juvenile features such as incomplete ossification of the nasals and the presence of short external gills. Impressions of the gill rakers are preserved in the gill chamber (Figs 1 and 2). These features indicate that the holotype was probably a neotenic form like its extant cryptobranchid relatives. Details of the body wall and tail are preserved. All of the bony tissues have been largely dissolved during fossilization; consequently, the skeleton is preserved as a natural mould that includes such details as sculptured surfaces on the parietals, squamosals, and the dorsal aspect of the pterygoids. The exceptional preservation of the specimen provides a variety of osteological and soft tissue characters to refer it to the Cryptobranchidae.

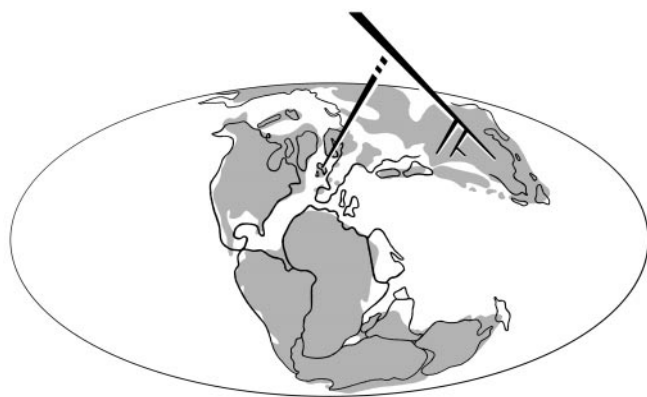
In addition, the new fossil assemblage includes a small larval specimen that shows remarkable preservation of soft tissues, with evidence of eyes, gill filaments, tail keel, tail seam, and notochord (Fig. 3). Also, there is incomplete ossification of the skull, vertebral

column and phalanges. Abundant intact conchostracans in the gut cavity indicate recent feeding behaviour for this individual. Although the lack of specific adult features makes it difficult to refer the larval form to the new taxon, the presence of uncapitate ribs supports its definitive inclusion in the Cryptobranchioidea.

Most phylogenetic hypotheses recognize that the Cryptobranchidae and the Hynobiidae form a monophyletic basal clade, the Cryptobranchioidea<sup>5,20–22</sup>. The split between the two groups is an important historical event in salamander evolution. Unfortunately, little is known about when and where this event took place. Until now, the earliest record of hynobiids was in the Miocene epoch (about 5.2 Myr) of Europe<sup>23</sup>, and that of cryptobranchids in the Palaeocene epoch (about 56 Myr) of both Asia and North America<sup>14–16</sup>. The presence of *Chunerpeton* in the Middle Jurassic of China implies that the split between hynobiids and cryptobranchids occurred before that time in Asia (Fig. 4). The hypothesis of an Asian origin of cryptobranchoids<sup>20</sup> is further supported by the Jurassic age and Asian distribution of stem caudates (such as *Karaurus*, *Kokartus*)<sup>24–26</sup> and other cryptobranchoid outgroups (for example, *Laccotriton*, *Sinerpeton*, *Jeholotriton*)<sup>5–7</sup>. The only exception to this palaeogeographic pattern is *Marmorerpeton*, a Jurassic stem caudate from Europe with uncertain affinities<sup>27</sup>. With an Asian origin of cryptobranchids, their extension to North



**Figure 3** A small larval cryptobranchoid showing exceptional preservation of soft tissues. The presence of uncapitate ribs supports its referral to the Cryptobranchioidea. con, conchostracans; eg, external gills; le, lens; nar, neural arch; nc, notochord; ph, phalanges.



**Figure 4** Map of Jurassic continental configuration (modified from ref. 29) with cladogram showing relationships and provenance of known Jurassic caudates. The taxa at the tips of the cladogram are, from left to right: (1) *Marmorierpeton*; (2) *Karaurus* and *Kokartus*; (3) *Laccotriton* and *Sinerpeton*; (4) *Jeholotriton*; and (5) *Chunerpeton*. The dotted line indicates the uncertainty of the phylogenetic position of *Marmorierpeton*. The shaded area represents the extent of continents.

America represents a more recent dispersal, perhaps associated with the development of a land bridge<sup>19,20</sup>. Hynobiids, on the other hand, underwent a range contraction. As hynobiid fossils are known from the Miocene and Pleistocene of Europe<sup>23</sup>, this group probably extended its range out of Asia only to be limited to that continent by the Holocene epoch.

Despite its Bathonian age, the new cryptobranchid shows extraordinary morphological similarity to its living relatives. This similarity underscores the stasis within salamander anatomical evolution<sup>28</sup>. Indeed, extant cryptobranchid salamanders can be regarded as living fossils whose structures have remained little changed for over 160 million years. Furthermore, the new material from China reveals that the early diversification of salamanders was well underway by the Middle Jurassic; several extant taxa including hynobiids and cryptobranchids had already appeared by that time. Notably, this ancient pattern of taxonomic diversification does not correlate to any great disparity in anatomical structure. □

Received 19 December 2002; accepted 3 February 2003; doi:10.1038/nature01491.

- Wake, D. B. & Larson, A. Multidimensional analysis of an evolving lineage. *Science* **238**, 42–48 (1987).
- Wake, D. B. Homoplasy: the result of natural selection or evidence of design limitations? *Am. Nat.* **138**, 543–567 (1991).
- Shubin, N., Wake, D. B. & Crawford, A. J. Morphological variation in the limbs of *Taricha granulosa* (Caudata: Salamandridae): evolutionary and phylogenetic implications. *Evolution* **49**, 874–884 (1995).
- Shubin, N. & Wake, D. B. Phylogeny, variation, and morphological integration. *Am. Zool.* **36**, 51–60 (1996).
- Gao, K.-Q. & Shubin, N. H. Late Jurassic salamanders from northern China. *Nature* **410**, 574–577 (2001).
- Gao, K., Cheng, Z. & Xu, X. First report of Mesozoic urodeles from China. *Geol. China* **248**, 40–41 (1998).
- Wang, Y. A new salamander (Amphibia: Caudata) from the Early Cretaceous Jehol Biota. *Vert. Palasiat.* **38**, 100–103 (2000).
- Dong, Z. & Wang, Y. A new urodele (*Liaoxitriton zhongjiani* gen. et sp. nov.) from the Early Cretaceous of western Liaoning Province, China. *Vert. Palasiat.* **36**, 159–172 (1998).
- Wang, S. et al. *Stratigraphical Lexicon of China: The Jurassic System* (Geological Publishing House, Beijing, 2000).
- Krzemiski, W. & Ren, D. *Praemacrolepis chinensis* n. sp. from the Middle Jurassic of China (Diptera: Tanyderidae). *Pol. Pis. Entomol.* **71**, 127–129 (2002).
- Ren, D. & Oswald, J. D. A new genus of kalligrammatid lacewings from the Middle Jurassic of China (Neuroptera: Kalligrammatidae). *Stuttgart. Beitr. Nat. B* **33**, 1–8 (2002).
- Ji, Q. & Yuan, C. Discovery of two kinds of protofeathered pterosaurs in the Mesozoic Daohugou Biota in the Ningcheng region and its stratigraphic and biologic significances. *Geol. Rev.* **48**, 221–224 (2002).
- Ren, D. et al. On the biostratigraphy of the Jurassic fossil beds at Daohugou near Ningcheng, Inner Mongolia. *Geol. Bull. China* **21**, 584–591 (2002).
- Naylor, B. G. Cryptobranchid salamanders from the Paleocene and Miocene of Saskatchewan. *Copeia* **1981**, 76–86 (1981).
- Estes, R. *Handbuch der Paläoherpetologie Teil 2 Gymnophiona, Caudata* (Gustav Fischer, Stuttgart, 1981).
- Gubin, Y. M. Paleocene salamanders from Southern Mongolia. *Paleontol. J.* **1991**, 91–102 (1991).
- Harland, B. W. et al. *A Geologic Time Scale 1989* (Cambridge Univ. Press, Cambridge, 1990).
- Evans, S. E. & Milner, A. R. A metamorphosed salamander from the Early Cretaceous of Las Hoyas, Spain. *Phil. Trans. R. Soc. Lond. B* **351**, 627–646 (1996).

- Milner, A. R. in *Amphibian Biology* Vol. 4 (eds Heatwole, H. & Carroll, R. L.) 1412–1444 (Surrey Beatty, Chipping Norton, Australia, 2000).
- Milner, A. R. in *Evolution, Time and Space: The Emergence of the Biosphere* (eds Sims, R. W., Price, J. H. & Whalley, P. E. S.) 431–468 (Academic, London, 1983).
- Duellman, W. E. & Trueb, L. *Biology of Amphibians* (McGraw-Hill, New York, 1986).
- Larson, A. & Dimmick, W. W. in *Amphibian Relationships: Phylogenetic Analysis of Morphology and Molecules* (eds Cannatella, D. & Hillis, D.) *Herpetol. Monogr.* **7**, 77–93 (1993).
- Venczel, M. & Hír, J. in *Abstracts of the Third World Congress of Herpetology* (eds Roček, Z. & Hart, S.) 218–219, (Third World Congress of Herpetology, Prague, 1997).
- Ivachnenko, M. F. Urodelans from the Triassic and Jurassic of Soviet Central Asia. *Paleontol. J.* **12**, 362–368 (1978).
- Nessov, L. A. Late Mesozoic amphibians and lizards of Soviet Middle Asia. *Acta Zool. Cracov.* **31**, 475–486 (1988).
- Nessov, L. A., Fedorov, P. V., Potanov, D. O. & Golovyeva, L. S. The structure of the skulls of caudate amphibians collected from the Jurassic of Kirgizstan and the Cretaceous of Uzbekistan. *Vest. Sankt-Peters. Univ. Geol.* **7**, 3–11 (1996).
- Evans, S. E., Milner, A. R. & Mussett, F. The earliest known salamanders (Amphibia: Caudata): A record from the Middle Jurassic of England. *Geobios* **21**, 539–552 (1988).
- Wake, D. B., Roth, G. & Wake, M. H. On the problem of stasis in organismal evolution. *J. Theor. Biol.* **101**, 211–224 (1983).
- Scotese, C. R. *Atlas of Earth History* Vol. 1 *Paleogeography* (Paleomap Project, Arlington, Texas, 2001).

**Acknowledgements** We thank Q. Ji for providing access to the holotype specimen. M. Ellison did the photographs of the holotype; K. Monoyios performed the drawings. This research was supported by the National Science Foundation, the National Natural Science Foundation of China (K.-Q.G.), and a grant from the John Simon Guggenheim Memorial Foundation (N.S.).

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to N.H.S. (e-mail: nshubin@uchicago.edu)

## Ecological and immunological determinants of influenza evolution

Neil M. Ferguson\*, Alison P. Galvani† & Robin M. Bush‡

\* Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, UK

† Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA

‡ Department of Ecology and Evolutionary Biology, University of California, Irvine, 92697, USA

In pandemic and epidemic forms, influenza causes substantial, sometimes catastrophic, morbidity and mortality. Intense selection from the host immune system drives antigenic change in influenza A and B, resulting in continuous replacement of circulating strains with new variants able to re-infect hosts immune to earlier types. This 'antigenic drift'<sup>1</sup> often requires a new vaccine to be formulated before each annual epidemic. However, given the high transmissibility and mutation rate of influenza, the constancy of genetic diversity within lineages over time is paradoxical. Another enigma is the replacement of existing strains during a global pandemic caused by 'antigenic shift'—the introduction of a new avian influenza A subtype into the human population<sup>1</sup>. Here we explore ecological and immunological factors underlying these patterns using a mathematical model capturing both realistic epidemiological dynamics and viral evolution at the sequence level. By matching model output to phylogenetic patterns seen in sequence data collected through global surveillance<sup>2</sup>, we find that short-lived strain-transcending immunity is essential to restrict viral diversity in the host population and thus to explain key aspects of drift and shift dynamics.

The surface glycoprotein haemagglutinin of influenza is under strong selection by the human immune system as the primary antibody target<sup>2</sup>. Figure 1a–c shows phylogenies constructed using



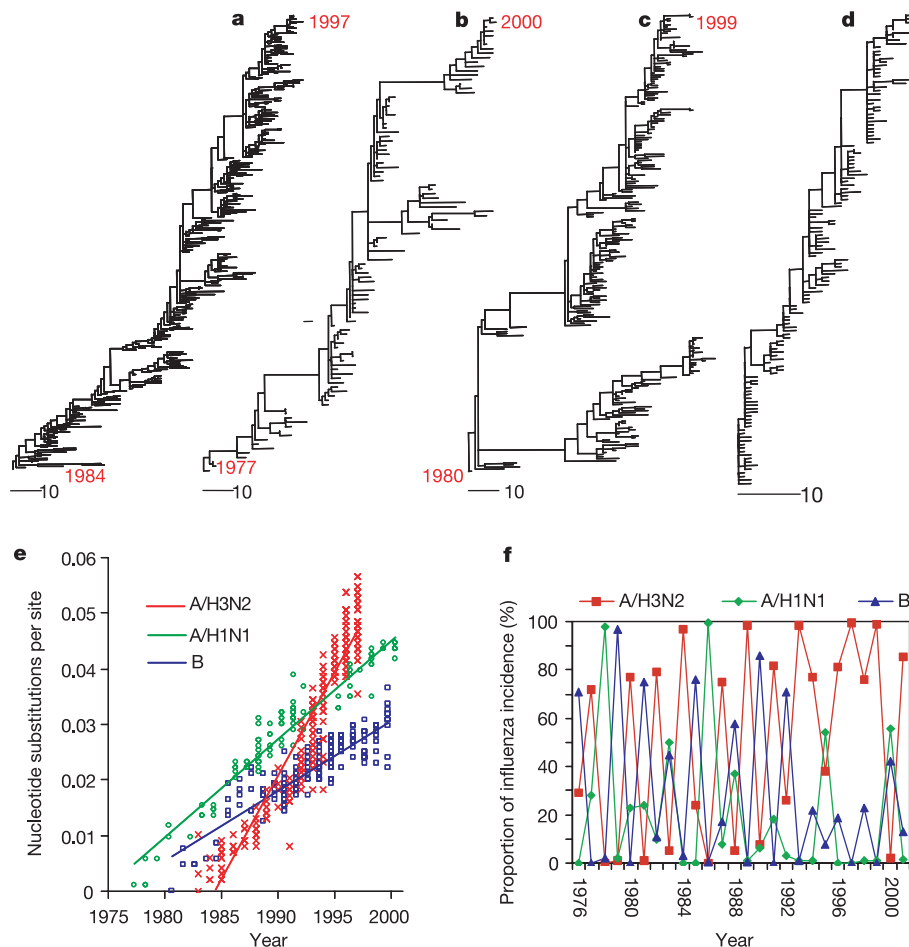
the HA1 domain of the haemagglutinin gene of human influenza A subtypes H3N2 (H3) and H1N1 (H1), and for influenza B. The slender 'trunk' of the H3 tree characterizes the evolutionary pattern of antigenic drift, and reflects the serial replacement of strains. Although brief co-circulation of closely related H3 lineages is not uncommon<sup>3</sup>, the H3 tree appears to be linear because the average survival time of side lineages is only 1.6 years (ref. 4). The H1 tree is similar to that of H3, although recent bifurcation of the main trunk has resulted in several extant lineages. It remains to be seen whether more than one will persist. Influenza B exhibits greater branching than either A subtype, having diverged into two major lineages in the early 1970s and multiple minor lineages thereafter<sup>5</sup>.

The explanation for the variation in tree shape between influenza viruses, illustrated in Fig. 1, is unknown. Furthermore, different strains evolve at different rates (Fig. 1e). Most studies suggest that H3 changes more rapidly than H1, while influenza B evolves more slowly than either A subtype<sup>3</sup>. The fixation rates calculated for the H3, H1 and B trees in Fig. 1a–c are consistent with this, at 0.0037, 0.0018 and 0.0013 nucleotide substitutions per site per year ( $\pm 0.001$ ), respectively. However, providing a biological explanation for these differences is complicated by immunological interactions between subtypes and types (Fig. 1f), although some of the

variation may solely be a function of the short time periods studied<sup>6</sup> (here 14, 24 and 20 years for H3, H1 and B respectively).

Fixation rates averaged over the whole of HA1 obscure variation among codons. For instance, 35% of replacements occur at 18 (5.5%) of the 329 codons in HA1 of H3. The excess of non-synonymous versus synonymous substitutions at these 18 codons reflects intense positive selection<sup>2,7</sup>, possibly because these codons are associated with the receptor binding site or antibody binding sites on the HA1 surface. The contrast between the fixation rate of 0.053 ( $\pm 0.005$ ) substitutions per site per year in the tree of the 18 key codons alone (Fig. 1d), and a rate of 0.0015 ( $\pm 0.0001$ ) for the 311 other codons of H3 reveals the importance of small subsets of codons to influenza evolution. We focus on identifying immunological and ecological processes that shape selection at antigenically significant codons to result in the observed evolutionary patterns of antigenic drift.

The diversity of a pathogen population is determined by the balance between the production of new strains and the competition-induced stochastic extinction of existing variants. To explain the surprisingly limited genetic diversity of influenza, we use an individual-based simulation (see Methods) to explore four determinants of this balance: the generation of new variants through mutation or reassortment, the initial establishment of those variants



**Figure 1** Evolutionary dynamics of influenza. **a**, Evolution of human influenza A subtype H3 from 1983 to 1997 derived from 357 sequences of 987 nucleotides<sup>2</sup>. **b**, Influenza A subtype H1, derived using 104 sequences of 1,032 nucleotides collected between 1977 and 2000. **c**, Influenza B, derived from 220 sequences of 1,041 nucleotides collected between 1980 and 1999. **d**, As **a**, but with horizontal distance representing the number of substitutions at the 18 positively selected codons<sup>2</sup> and with zero-length branches removed. All trees were generated from HA1 sequences using the tree bisection-reconnection branch-swapping option of the heuristic search option of the maximum-

parsimony routine of PAUP<sup>29</sup>. Tree shape is robust to other phylogenetic algorithms. (See Supplementary Information for sequence data accession numbers.) **e**, Distance from the root of the tree to each terminal node in nucleotide substitutions against year of isolation. Slopes of lines illustrate differences in fixation rates between H3, H1 and B. **f**, Proportion of laboratory-analysed influenza infections isolates identified as H3, H1 or B each year by the World Health Organization laboratories in the United States from 1976 to 2001 (ref. 30). The trends shown are suggestive of interference between subtypes, with annually fluctuating subtype dominance.

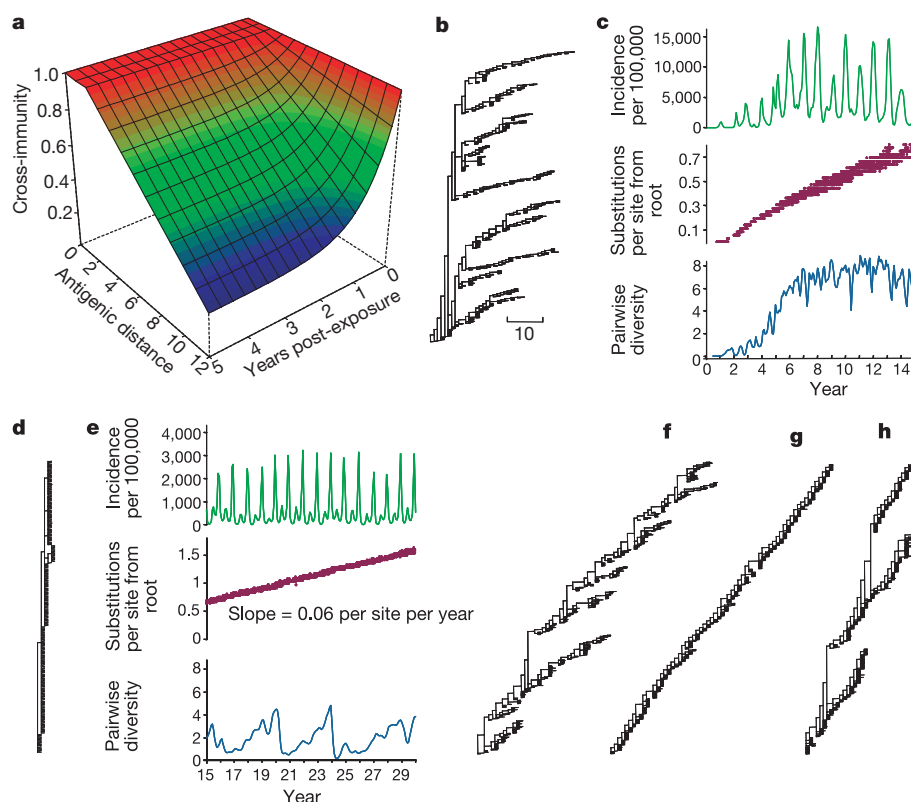
in the host population, cross-immunity-mediated competition between strains, and the processes governing the rate of strain extinction. Although our model is computationally intensive, it captures realistic population genetics of large numbers of unique co-circulating strains, a simplified representation of the spatial structure of human populations and contact patterns, and the stochastic nature of mutation and extinction. In addition, it enables us to compare model output with both phylogenetic and epidemiological data.

We focus on the role of immunity in imposing density-dependent constraints that limit influenza diversity. Influenza antibody recognition tends to decrease with genetic divergence of HA<sub>1</sub>, indicating that cross-protection arising from long-lived memory responses decreases as the sequence divergence between two strains increases (Fig. 2a). In such circumstances, as the antigenic diversity of co-circulating strains increases, so should their mean infection prevalence, thereby increasing production of further variants. If such cross-immunity is the only form of interaction between strains, the expected result of this selection for antigenic novelty is exponential growth of diversity, with concomitant excessive infection incidence as shown in Fig. 2b, c. This diversification can be mitigated by lowering mutation rates and increasing the intensity of cross-immunity (Fig. 2d)—but only at the cost of greatly slowing the

rate of evolution, resulting in unrealistically low fixation probabilities and diversity levels.

More extensive sensitivity analysis, described in detail in the Supplementary Information, suggests that it is impossible to reproduce the observed limited diversity and high fixation rates of influenza A and B without an additional, nonspecific competitive interaction between strains. In particular, we examined the effect of functional constraints on viral evolution in three ways: first, as a crude representation of changes occurring in parts of the viral genome that were not being explicitly modelled, we assumed that every new variant had a random fitness (that is, transmissibility). Second, we explicitly modelled a set of codons that determine viral transmissibility independently from those determining antigenicity. Third, we assumed that there was a transmissibility gradient associated with antigenic space. In each case functional constraints alone are insufficient to explain the limited diversity of influenza.

The primary candidate for a nonspecific competitive interaction between strains is a second immune-response component: short-lived immunity that decays rapidly with time from last exposure and inhibits reinfection by any new strain. Such a response acts as a density-dependent constraint on overall infection incidence, thus limiting effective viral population size and reducing the likelihood of explosive diversity growth.

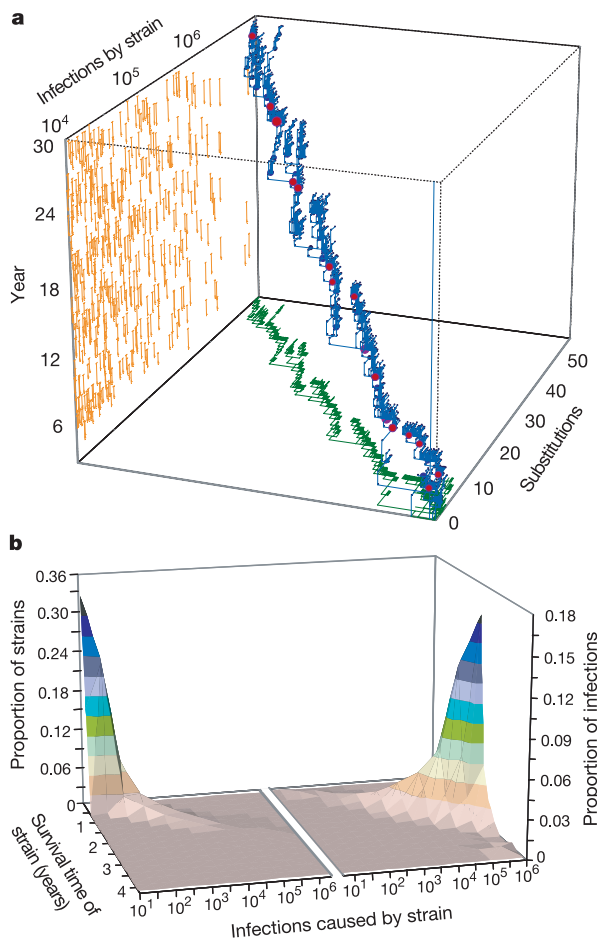


**Figure 2** Model drift dynamics for a single subtype. **a**, Cross-immunity profile assumed for baseline model parameters (see Methods), composed of a long-lived strain-specific memory response and a short-lived nonspecific response. The former gives 99% cross-immunity ( $\theta_1 = 0.99$ ) for strains up to an escape threshold of  $n_t = 2$  changes apart, which drops linearly with increasing genetic distance to a level of  $\theta_0 = 0.25$  for strains with differences at all 12 codons. The latter gives complete protection against reinfection immediately after an infection, but this protection decays exponentially over time with a half-life of six months. **b**, 50-year model tree for baseline parameters but no short-lived nonspecific immunity and mutation rate  $\delta = 10^{-6}$ , showing explosive diversification. For clarity, in every model tree only the 500 strains causing most infections are shown. **c**, 15 years output from model run shown in **b** showing time series of weekly infection incidence in a single geographical patch (upper), the number of nucleotide substitutions

per site (relative to the original strain) for new strains arising through time (middle), and the prevalence-weighted mean pairwise number of amino acid differences between all co-circulating strains (lower)—a measure of diversity. Genetic bottlenecks due to seasonal forcing eventually cause diversity to saturate. **d**, Without short-lived immunity, lowering mutation rate ( $\delta = 2.5 \times 10^{-6}$ ) and/or increasing the immune escape threshold ( $n_t = 4$ ) is only able to achieve constant diversity at the cost of very low fixation rates. **e**, As **c**, but for a model with baseline parameters including short-lived immunity. **f**, 50-year tree from model with baseline parameters including short-lived immunity. **g**, Over-restricted diversity achieved with  $\delta = 10^{-6}$  and  $n_t = 1$ . **h**, Reducing the intensity of cross-immunity (maximum level  $\theta_1 = 0.8$ , minimum level  $\theta_0 = 0.5$ ,  $n_t = 1$ ) and the mutation rate ( $\delta = 2.5 \times 10^{-6}$ ) lowers fixation rate and gives a tree more reminiscent of influenza B.

With the resulting two-component immune memory response (Fig. 2a), our model reproduces key aspects of the dynamics of influenza epidemiology and evolution (Fig. 2e), including the magnitude and seasonality of influenza incidence, a fairly stable viral fixation rate through time<sup>4</sup>, and temporal fluctuations in viral diversity<sup>8</sup> driven by the generation of short-lived viral sublineages and their eventual extinction over the course of several years. Our simulated phylogeny (Fig. 2f) is very similar to the actual pattern of H3 evolution (Fig. 1d). Observed fixation rates can be reproduced at a variety of parameter values (Fig. 2g) by trading off cross-immunity intensity against mutation rates. However, matching both fixation rates and detailed tree shape significantly restricts the space of feasible parameters. Achieving a lower fixation rate and higher level of diversity, such that the resulting tree (Fig. 2h) is more akin to influenza B (Fig. 1c) than to influenza A H3 (Fig. 1a), requires reducing both the intensity of cross-immunity and rate of mutation.

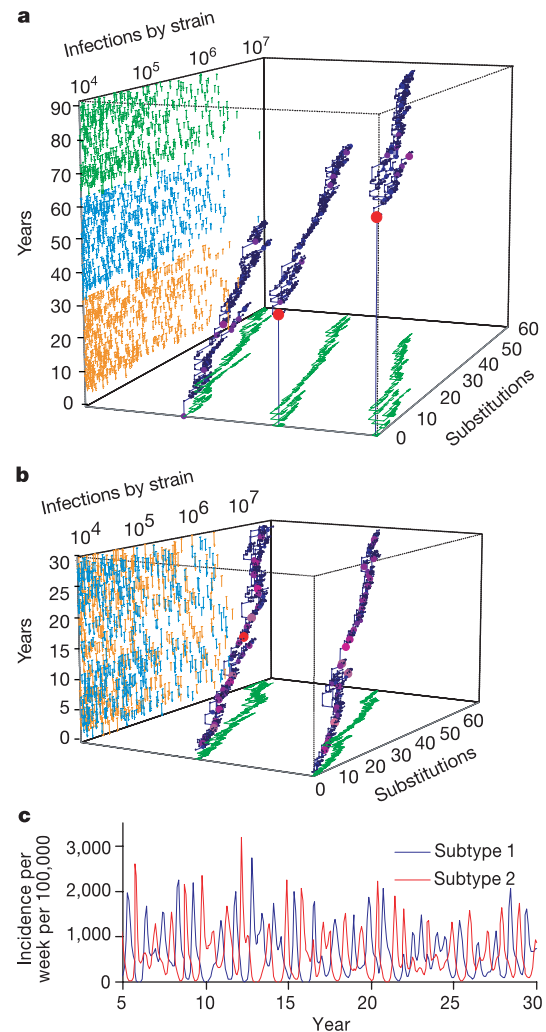
One of the key advantages of our modelling approach is that complete information is provided on the temporal dynamics of both evolution and transmission of all virus variants—information that is difficult to collect from surveillance. Figure 3a illustrates this



**Figure 3** Detailed evolutionary dynamics of model with short-lived nonspecific immunity. **a**, Three-dimensional tree with genetic distance and time of strain emergence along the y and z axes respectively. The size and colour of internal nodes represent the infection incidence caused by that strain. The traditional two-dimensional phylogenetic tree is projected on the x-z plane. On the y-z plane, each strain causing over 10<sup>4</sup> infections is represented by a line, the start and end of which (along the z axis) indicate time of creation and extinction, and the y-position indicates the total incidence due to that strain. **b**, Joint distribution of strain survival time and log<sub>10</sub> infections caused by each strain. The left pane shows the probability density function calculated for all strains generating over ten infections (averaged over five model realizations), and the right pane the density function for all strains weighted by the number of infections they cause.

point by showing the correlation between genetic distance and time in sequence evolution, and the variability in frequency of different strains. The phylogeny of influenza is seen to arise from multi-strain population dynamics characterized by a highly skewed survival-time distribution for strains (Fig. 3b), where 95% of strains survive for less than a year and infect fewer than 100 people. The less than 1% of strains that successfully become globally established account for 90% of all disease incidence while surviving for 2–3 years. This timescale is consistent with observed survival times of new dominant drift variants<sup>4</sup>.

Two pandemics have occurred since the influenza virus was



**Figure 4** Model dynamics for interacting subtypes. **a**, Single realization of model with baseline parameters, introducing new subtypes as strains with random genotypes into a single host in the population in years 34 and 64 of a 92-year simulation. Plot shown as Fig. 2c, but three distinct trees can be seen, corresponding to the three pandemic lineages. The strain scatter-plot on the y-z plane is coloured according to subtype. Subtype replacement during pandemics occurred in 100% of the total of 25 model realizations examined. **b**, Subtype coexistence can also occur for these parameter values if the proportions of the host population with immunity to each subtype are close to equilibrium values. The plot shows two subtypes co-evolved from the two-strain equilibrium. **c**, The resulting global incidence time series for co-evolving subtypes predicts that the immunological competition generated by short-lived subtype-transcending immunity would be observable through partial competitive exclusion of different subtypes in sequential years. For **a–c**,  $\tau_{1/2} = 0.1$  years for between-subtype protection and 0.5 years for within-subtype protection. Subtype coexistence is possible, but is an unlikely outcome, if  $\tau_{1/2} = 0.5$  for both within- and between-subtype protection. However, for larger host population sizes, coexistence may become more likely even for that parameter choice.



isolated in 1930. In each case (1957 and 1968) the pandemic strain, an avian–human influenza A reassortant, replaced the previously circulating strain of influenza A. However, coexistence is possible (Fig. 4b) in situations where subtype introduction does not generate a pandemic, as indicated by the current co-circulation of H3 and H1 after the reintroduction of H1 in 1977 into a population with considerable residual herd-immunity against this subtype. The proposed two-component immune response also helps to explain these patterns of competitive exclusion and coexistence between subtypes. Short-lived nonspecific (heterosubtypic) immunity that inhibits immediate reinfection is critical to capture the dynamics of subtype replacement during a pandemic (Fig. 4a).

In the absence of heterosubtypic immunity, the antigenic distance between subtypes generates very limited inter-subtype competition, and subtype elimination becomes unfeasible. With such immunity, subtype extinction becomes highly probable in the context of the globally synchronized large-scale transient dynamics associated with pandemics. However, where subtype introduction causes a smaller epidemic (for example, because of some pre-existing immunity to the new subtype) coexistence is possible if the duration (or effectiveness) of nonspecific immunity is greater within subtypes than between subtypes, a conclusion for which there is abundant empirical evidence<sup>9</sup>. Such immunity gives the less prevalent subtype a frequency-dependent advantage, mitigating the highly unstable dynamics which otherwise rapidly drive the rarer subtype extinct. The resulting incidence time series shows a pattern of fluctuating partial competitive exclusion between subtypes (Fig. 4c), consistent with epidemiological surveillance data (Fig. 1f).

Transient nonspecific immunity is paramount in the generation of such dynamics, and in limiting diversity within subtypes—although other mechanisms that induce nonspecific inter-strain competition (for example, disease-associated reductions in host contact rates) may also contribute. Persistent nonspecific immunity of lower intensity rapidly saturates to affect nearly all hosts and hence fails to impose the intense density-dependent competition that limits viral diversity.

Viral pathogens give unique insight into evolution because their short lifecycle allows dynamical data to be collected at both ecological and evolutionary timescales. Here we show that mathematical models can now give an integrated and detailed description of processes coupling these timescales, reproducing both epidemiological dynamics and patterns of sequence evolution. Thus the influence of epidemiological processes and immunological phenotype on emergent evolutionary dynamics can be explored at a level of realism not previously feasible. Quantitative information on the magnitude of factors determining evolutionary patterns can also be derived (for example, the relationship between *in vitro* haemagglutination assays and the protection afforded against infection with heterologous strains) that would otherwise involve epidemiological and clinical studies of impractically large scale.

Antigenic drift in human influenza has been attributed to step-wise selection for antigenic novelty resulting from an immunity phenotype in which cross-protection declines as the genetic distance between strains increases<sup>10,11</sup>. However, our study shows that without severe density-dependent constraints on the effective population size of the virus, this process alone would lead to rapidly increasing diversity of viral populations. Within a wide range of model variants and parameter ranges explored, transient nonspecific immunity that either prevents rapid reinfection or reduces transmissibility was the only clear candidate for such a constraint. A number of studies support the existence of short-lived nonspecific within-subtype<sup>10–12</sup> and between-subtype<sup>13,14</sup> immune responses, with CTL<sup>15</sup>, natural killer cells<sup>16</sup> and cytokines being proposed effector mechanisms<sup>17</sup>. Quantitative data on the duration of such immunity and its effect on susceptibility or infectiousness is limited, but the timescales of weeks to months explored in this paper are consistent with epidemiological and experimental studies<sup>14,16,18,19</sup>.

Our analysis gives some insight into the potential causes of differences between H1, H3 and B (Fig. 1a–c), as detailed more fully in the Supplementary Information. Alone, the lower fixation rate of H1 and B might simply reflect a lower net mutation rate. However, variation in the levels of genetic diversity between H1, H3 and B may indicate differences in cross-immunity phenotype; that is, reduced maximal cross-immunity, or a lower threshold number of amino acid changes required for evasion of pre-existing immunity. Fuller explanation of differences between influenza viruses circulating within human and other mammalian species<sup>20</sup> will require rigorous analysis of the key loci involved in antigenic selection, together with a more complete model of subtype coevolution. With robust parameterization of population demography and structure, modelling should also give insight into whether the origins of apparent differences are largely ecological (for example, lower population contact rates) or reflect underlying variation in viral phenotype.

Limitations of available quantitative data require models to simplify several aspects of influenza biology and epidemiology<sup>21</sup>. More complete descriptions of antigenicity and other aspects of viral phenotype (such as virulence or transmissibility) will further improve understanding of the constraints on viral evolution. Coupled with more sophisticated representations of spatial heterogeneity in host density and human contact patterns, model correspondence with epidemiological and sequence data will also be enhanced. Collection of such data depends on continued improvement in global surveillance of new influenza strains<sup>22</sup>. Current programmes successfully identify new variants, but the quantitative tracking of strain-specific incidence—although increasingly feasible with the advent of rapid molecular typing methods—remains a long-term objective. This work is essential if statistically predictive modelling is to be achieved, whether applied to predicting new vaccine candidate strains<sup>3,23</sup>, anticipating the effect of increasing population size and vaccine use on viral evolution, or modelling the global spread of future pandemics. □

## Methods

### Transmission model

Baseline parameter values (selected from the detailed sensitivity analyses described in the Supplementary Information as best reproducing observed patterns of evolution) are given here; parameters used in sensitivity analyses are detailed in the text. We simulate a constant population size of  $N$  ( $=12$  million) hosts, each with lifespan  $L$  ( $=30$  years), equally distributed between  $M$  ( $=20$ ) geographical patches. The locations of hosts in a patch are randomly distributed with a mean separation of one (in arbitrary units). Each host is labelled by patch index  $p$  ( $1 \leq p \leq M$ ) and host index  $i$  ( $1 \leq i \leq N/M$ ), and at any time is characterized by four variables,  $\{B_{p,i}, T_{p,i}, S_{p,i}, H_{p,i}\}$ , where  $B_{p,i}$  is the time of birth,  $T_{p,i}$  is the time of last infection,  $S_{p,i}$  is the index of the last strain with which the host was infected, and  $H_{p,i}$  characterizes the immune history of the host. Where no prior infection has occurred,  $S = 0$  and  $T = -\infty$ . On infection, hosts are assumed to incubate infection for 2 days (ref. 24), be infectious for the following 4 days (ref. 25), and recover thereafter. Results here assume no co-infection, but are not sensitive to this assumption. Infection and mutation processes were modelled stochastically using a discrete-time model formulation with a one-day time-step. On day  $t$ , the probability a host is exposed to strain  $s$  is given by

$$\rho_{p,i}(t,s) = 1 - \exp \left\{ - \left[ 1 + \varepsilon_p \sin(t/365) \right] \left( \sum_{\substack{i' \in \{S_{p,i'}=s\} \cap \\ \{2 < t - T_{p,i'} \leq 6\} \cap \\ \{r_p(i,i') \leq R\}}} \beta_L + \sum_{\substack{i' \in \{S_{p,i'}=s\} \cap \\ \{2 < t - T_{p,i'} \leq 6\}}} \beta_W \right) - \sum_{p' \neq p} \left[ 1 + \frac{1}{2}(\varepsilon_p + \varepsilon_{p'}) \sin(t/365) \right] \sum_{\substack{i' \in \{S_{p',i'}=s\} \cap \\ \{2 < t - T_{p',i'} \leq 6\}}} \beta_B \right] \right\}$$

where  $r_p(i,i')$  is the distance between hosts  $i$  and  $i'$  in patch  $p$ ,  $R$  ( $=4$ ) is the threshold distance for local transmission,  $\beta_L$  is the contact rate between hosts in the same spatial locality (defined by  $r_p(i,i') \leq R$ ),  $\beta_W$  is a contact rate representing a less frequent process of mixing between all hosts in a patch, and  $\beta_B$  is the rate of between-patch contacts.  $\beta_B$  and

$\beta_L$  are assigned values to give a basic reproduction number,  $R_0$ , for local transmission in a patch of 5,  $R_0$  for mass-action transmission in a patch of 0.4, and an  $R_0$  of 0.02 between any two patches. Seasonal variation in contact rates is characterized by  $\epsilon_p$ , and is assumed to be opposite in phase for the Northern and Southern hemispheres (represented by setting  $\epsilon_p = +0.25$  for  $1 < p \leq M/2$ , and  $\epsilon_p = -0.25$  for  $M/2 < p \leq M$ ).

The probability that a host is infected upon exposure depends on the strain  $s$ , and the host's immune history. Each strain is characterized by  $A$  epitopes, each consisting of  $C$  codons (three nucleotide bases). Immunity is assumed to be specific to the set of amino acids to which the host has been exposed at each codon, and for simplicity, no functional constraints are imposed on the amino acid sequences. For the default values of  $A = 4$  and  $C = 3$  used here, a total of  $4 \times 10^{15}$  strains are possible. Antigenic distance,  $d(s, H)$  between a strain  $s$  and the immune history of a host,  $H$ , is then simply defined as the number of codons in strain  $s$  for which the amino acid has not been previously encountered by the host. The level of cross-protection from infection provided at a certain antigenic distance is given by the function  $f(d)$ , where we assume  $f(d) = \theta_1 + (\theta_0 - \theta_1)(d - n_t)/(AC - n_t)$  for  $d \geq n_t$ ,  $f(d) = \theta_1$  for  $0 < d < n_t$ , and  $f(d) = 1$  for  $d = 0$ .  $n_t (=2)$  is the threshold level of change necessary for cross-protection to drop below the maximal level set by  $\theta_1 (=0.99)$ , and  $\theta_0 (=0.25)$  is the minimal level of cross-protection mounted against a strain with no similarity to a previously encountered strain at the codons modelled ( $=0$  for no cross-protection). These assumptions reflect empirical studies suggesting that two or more substitutions at key antigenic sites are required to escape pre-existing immunity<sup>26,27</sup>.

The probability that a host will be infected by a strain following exposure is given by

$$\varphi_{p,i}(t, s) = \{1 - \omega \exp[-(t - T_{p,i})/\tau]\} [1 - f(d(s, H_{p,i}))]$$

where  $\tau$  ( $\approx 270$  days) is the decay timescale (half-life =  $\tau_{1/2} = 187$  days  $\approx 6$  months) and  $\omega$  ( $\approx 1$ ) is the peak level ( $0 \leq \omega \leq 1$ ) of a short-lived strain-transcending immune response that protects against reinfection in the weeks after an infection (see main text). If a host is infected following exposure, its immune history is updated to include the new strain. If exposure does not result in infection, we assume that no immune response is raised to the new strain but that pre-existing immune responses are boosted (akin to the 'original antigenic sin' response<sup>28</sup>), by resetting  $T_{p,i}$  to  $t - 6$  immediately after exposure.

There is a probability  $\delta$  ( $\approx 10^{-5}$ ) per base per day that a nucleotide substitution will occur in the virus in an infected host and the resulting mutant strain will replace the pre-existing strain in that individual. The individual is then infectious with the new strain. All strains are assumed to have the same intrinsic transmissibility (but see Supplementary Information), and model runs were started from near the single-strain equilibrium.

Received 18 November 2002; accepted 21 February 2003; doi:10.1038/nature01509.

- Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M. & Kawoka, Y. Evolution and ecology of influenza A viruses. *Microbiol. Rev.* **56**, 152–179 (1992).
- Bush, R. M., Fitch, W. M., Bender, C. A. & Cox, N. J. Positive selection on the H3 hemagglutinin gene of human influenza virus A. *Mol. Biol. Evol.* **16**, 1457–1465 (1999).
- Cox, N. et al. in *Options for the Control of Influenza II* (eds Hannoun, C., Kendal, A. P., Klenk, H. D. & Ruben, F. L.) 223–230 (Elsevier Science, Amsterdam, 1993).
- Fitch, W. M., Bush, R. M., Bender, C. A. & Cox, N. J. Long term trends in the evolution of H(3) HA1 human influenza type A. *Proc. Natl Acad. Sci. USA* **94**, 7712–7718 (1997).
- Lindstrom, S. E. et al. Comparative analysis of evolutionary mechanisms of the hemagglutinin and three internal protein genes of influenza B virus: multiple cocirculating lineages and frequent reassortment of the NP, M, and NS genes. *J. Virol.* **73**, 4413–4426 (1999).
- Daly, J., Wood, J. & Robertson, J. in *Textbook of Influenza* (eds Nicholson, K., Webster, R. & Hay, A.) 168–177 (Blackwell Science, Oxford, 1998).
- Bush, R. M., Bender, C. A., Subbarao, K., Cox, N. J. & Fitch, W. M. Predicting the evolution of human influenza A. *Science* **286**, 1921–1925 (1999).
- Cox, N. & Regnery, H. in *Options for the Control of Influenza III* (eds Brown, L. E., Hampson, Q. W. & Webster, R. G.) 591–598 (Elsevier Science, Amsterdam, 1996).
- Glezen, W. P. & Couch, R. B. in *Viral Infections of Humans* (eds Evans, A. S. & Kaslow, R. A.) 473–505 (Plenum Medical, New York, 1997).
- Gill, P. W. & Murphy, A. M. Naturally acquired immunity to influenza type A. A further prospective study. *Med. J. Aust.* **2**, 761–765 (1977).
- Frank, A. L., Taber, L. H. & Porter, C. M. Influenza B virus reinfection. *Am. J. Epidemiol.* **125**, 576–586 (1987).
- Frank, A. L. & Taber, L. H. Variation in frequency of natural reinfection with influenza A viruses. *J. Med. Virol.* **12**, 17–23 (1983).
- Yetter, R. A., Lehrer, S., Ramphal, R. & Small, P. A. Outcome of influenza infection: effect of site of initial infection and heterotypic immunity. *Infect. Immun.* **29**, 654–662 (1980).
- Sonoguchi, T., Naito, H., Hara, M., Takeuchi, Y. & Fukumi, H. Cross-subtype protection in humans during sequential, overlapping, and/or concurrent epidemics caused by H3N2 and H1N1 influenza viruses. *J. Infect. Dis.* **151**, 81–88 (1985).
- Yewdell, J. W., Bennink, J., Smith, G. L. & Moss, B. Influenza A virus nucleoprotein is a major target antigen for crossreactive anti-influenza A virus cytotoxic T lymphocytes. *Proc. Natl Acad. Sci. USA* **82**, 1785–1789 (1985).
- Skoner, D. P. et al. Effect of influenza A virus infection on natural and adaptive cellular immunity. *Clin. Immunol. Immunopathol.* **79**, 294–302 (1996).
- Seo, S. H., Peiris, M. & Webster, R. G. Protective cross-reactive cellular immunity to lethal A/Goose/Guangdong/1/96-like H5N1 influenza virus is correlated with the proportion of pulmonary CD8(+) T cells expressing gamma interferon. *J. Virol.* **76**, 4886–4890 (2002).
- Samhara, S. et al. Heterosubtypic immunity against human influenza A viruses, including recently emerged avian H5 and H9, induced by FLU-ISCOM vaccine in mice requires both cytotoxic T-lymphocyte and macrophage function. *Cell Immunol.* **211**, 143–153 (2001).
- McElhaney, J. E., Meneilly, G. S., Pinkoski, M. J., Lechelt, K. E. & Blackley, R. C. Vaccine-related determinants of the interleukin-2 response to influenza vaccination in healthy young and elderly adults. *Vaccine* **13**, 6–10 (1995).
- Nicholson, K. G., Webster, R. G. & Hay, A. J. *Textbook of Influenza* (Blackwell Science, Oxford, 1998).

- Earn, D. J., Dushoff, J. & Levin, S. A. Ecology and evolution of the flu. *Trends Ecol. Evol.* **17**, 334–340 (2002).
- Layne, S. P. et al. A global lab against influenza. *Science* **293**, 1729 (2001).
- Bush, R. M., Bender, C. A., Subbarao, K., Cox, N. J. & Fitch, W. M. Predicting the evolution of human influenza A. *Science* **286**, 1921–1925 (1999).
- Moser, M. R. et al. An outbreak of influenza aboard a commercial airliner. *Am. J. Epidemiol.* **110**, 1–6 (1979).
- Benenson, A. S. *Control of Communicable Diseases in Man* (American Public Health Association, Washington DC, 1975).
- Wilson, I. A. & Cox, N. J. Structural basis of immune recognition of influenza virus hemagglutinin. *Annu. Rev. Immunol.* **8**, 737–771 (1990).
- Smith, C. B., Cox, N. J., Subbarao, K., Taber, L. H. & Glezen, W. P. Molecular epidemiology of influenza A(H3N2) virus reinfections. *J. Infect. Dis.* **185**, 980–985 (2002).
- Hoskins, T. W., Davies, J. R., Smith, A. J., Miller, C. L. & Allchin, A. Assessment of inactivated influenza-A vaccine after three outbreaks of influenza A at Christ's Hospital. *Lancet* **1**, 33–35 (1979).
- Swofford, D. L. *PAUP\*: Phylogenetic Analysis Using Parsimony* (Sinauer, Sunderland, 1998).
- Morbidity and Mortality Weekly Reports (Centers for Disease Control and Prevention, annual reports on influenza frequency, 1986–2002).

**Supplementary Information** accompanies the paper on Nature's website (<http://www.nature.com/nature>).

**Acknowledgements** We thank N. J. Cox for discussions. N.M.F. thanks the Royal Society, Howard Hughes Medical Institute and Medical Research Council, A.P.G. thanks the Miller Institute, and R.M.B. thanks the NIH for research funding.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to N.M.F. (e-mail: neil.ferguson@ic.ac.uk).

## Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events

John E. Bowers\*, Brad A. Chapman\*, Junkang Rong & Andrew H. Paterson

Plant Genome Mapping Laboratory, University of Georgia, Athens, Georgia 30602, USA

\* These authors contributed equally to this work

Conservation of gene order in vertebrates is evident after hundreds of millions of years of divergence<sup>1,2</sup>, but comparisons of the *Arabidopsis thaliana* sequence<sup>3</sup> to partial gene orders of other angiosperms (flowering plants) sharing common ancestry  $\sim 170$ –235 million years ago<sup>4</sup> yield conflicting results<sup>5–11</sup>. This difference may be largely due to the propensity of angiosperms to undergo chromosomal duplication ('polyploidization') and subsequent gene loss<sup>12</sup> ('diploidization'); these evolutionary mechanisms have profound consequences for comparative biology. Here we integrate a phylogenetic approach (relating chromosomal duplications to the tree of life) with a genomic approach (mitigating information lost to diploidization) to show that a genome-wide duplication<sup>3,13–17</sup> post-dates the divergence of *Arabidopsis* from most dicots. We also show that an inferred ancestral gene order for *Arabidopsis* reveals more synteny with other dicots (exemplified by cotton), and that additional, more ancient duplication events affect more distant taxonomic comparisons. By using partial sequence data for many diverse taxa to better relate the evolutionary history of completely sequenced genomes to the tree of life, we foster comparative approaches to the study of genome organization, consequences of polyploidy, and the molecular basis of quantitative traits.

Angiosperms sustain humanity by providing oxygen, medicines, food, feed, fibre, fuel, erosion and flooding control, soil regener-

ation and other benefits<sup>18</sup>, warranting increased exploration of their genomic diversity. The ~1,000-fold variation in angiosperm genome size, from 125 million base pairs (Mbp) for *Arabidopsis thaliana*<sup>3</sup> to 124,852 Mbp for *Fritillaria assyriaca*<sup>19</sup>, motivates comparative approaches, using data from smaller genomes to accelerate the study of larger genomes. It was recognized recently that even small angiosperm genomes contain much duplication<sup>3,13–17</sup>, but robust application of this finding to comparative biology has awaited a means to directly relate chromosomal events to the tree of life in a manner that is not subject to the differing rates of various molecular clocks<sup>20,21</sup>.

After revising our earlier analysis of *Arabidopsis thaliana* chromosomal duplication<sup>13</sup> to reflect matches between inferred protein (instead of DNA) sequences and to include 26,028 (instead of 15,199) genes (obtained from NCBI), we circumscribed 34 non-overlapping chromosomal segment pairs that include 23,177 *Arabidopsis* genes (representing 89% of the total) (Fig. 1a). Circumscription of these segment pairs (here called the  $\alpha$  group) is conservative, as  $\chi^2$  tests comparing the observed number of gene duplications comprising each pair to the number expected in a chromosomal region of equal size if duplications are randomly distributed over the genome, show a maximal random likelihood of  $2.8 \times 10^{-84}$  (for  $\alpha$ 17). Of 2,851 (11%) genes in putatively non-duplicated regions, 1,570 (55%) were near centromeres, and the rest in gaps between the duplications.

After using interpolation to infer an ancestral gene order (see Supplementary Information) that accounts for the composition of the 26 'large' (Fig. 1 legend)  $\alpha$  segment pairs that collectively comprise 83% of the transcriptome, a second iteration of analysis revealed additional, more ancient duplications. Nested within the 26  $\alpha$  segment pairs, we circumscribed 29 additional duplications (Fig. 1b). All 29 segments are mosaics of genes from each of their four 'descendant' modern chromosome segments, and most are only evident by analysis of the inferred ancestral gene order. The 29 segment pairs comprise two subpopulations (called here  $\beta$  and  $\gamma$ ).  $\beta$ 01– $\beta$ 22 include 13,449 genes (51.6% of the transcriptome) and are non-overlapping, with a maximal probability of random occurrence of  $3 \times 10^{-3}$  (for  $\beta$ 19).  $\beta$ 18, 20 and 22 each comprise <1% of the transcriptome, and like the 'small'  $\alpha$  pairs were grouped for further analyses.  $\gamma$ 01– $\gamma$ 07 include 5,287 genes (20.3% of the transcriptome) with a maximal probability of random occurrence of  $<2 \times 10^{-3}$  (for  $\gamma$ 04), and overlap with many  $\beta$  segments. Distinct identities of the  $\beta$  and  $\gamma$  groups are reinforced by the finding that among 78 cases in which protein matches were found in both groups, 59 (75%) showed greater distance between the members of the  $\gamma$  pairs.

To relate the events that produced the  $\alpha$ ,  $\beta$  and  $\gamma$  segment pairs to the angiosperm family tree, we compared all syntenic *Arabidopsis* gene pairs from each duplication event (respectively), to individual genes from representatives of the gymnosperms (Pinaceae), monocotyledonous angiosperms (*Oryza*), asterids (Solanaceae), distantly related rosids (*Glycine*, *Medicago*), closely related rosids (Malvaceae) and a confamilial Brassicaceae genus (*Brassica*), respectively (see Fig. 2 for GenBank taxon IDs). Although the ideal comparison would involve analysis of individual gene trees that included full-length sequences for all taxa simultaneously, the fragmentary nature of expressed sequence tags (ESTs) that comprise most plant gene databases necessitated a modified approach. First, using genes for which adequate sequence information existed (based on criteria described in Fig. 2 and Methods), we made phylogenetically rooted trees using *Physcomitrella* (a moss) as an outgroup, then evaluated the frequency at which heterologous proteins clustered internally or externally to the *Arabidopsis* duplicates. For many additional genes, it was possible to determine whether inferred protein sequences from the syntenic *Arabidopsis* genes were more, or less, similar to one another than to the heterologous protein by evaluating PAM (per cent accepted mutation)-based pairwise distances (Fig. 2 and

Methods). Both analyses considered only matches of  $\geq 35$  amino acids (105 nucleotides). (Lists of genes and their arrangements in the  $\alpha$ ,  $\beta$  and  $\gamma$  duplications, inferred ancestral gene orders, frequencies of internal versus external rooted trees and PAM-based pairwise distances for each  $\alpha$ ,  $\beta$  and  $\gamma$  segment pair, and the accession numbers plus chromosomal and map locations of the cotton DNA sequences used for synteny analysis are available as Supplementary Information.)

The *Arabidopsis*  $\alpha$  duplication event pre-dated its divergence from *Brassica* about 14.5–20.4 million years (Myr) ago<sup>4</sup>, but post-dated its divergence from the Malvaceae 83–86 Myr ago<sup>22</sup>. Rooted trees and PAM-based pairwise distances (Fig. 2b) showed that 49% and 64% (respectively) of *Brassica* sequences were more similar to one duplicated *Arabidopsis* sequence than was the other *Arabidopsis* sequence. By contrast, only 6–19% of Malvaceae and more distant sequences clustered internally to the *Arabidopsis* syntenic duplicates.

The  $\beta$  event pre-dated *Arabidopsis* divergence from the other dicots studied, but post-dated divergence from the monocots about 170–235 Myr ago<sup>4</sup>. Rooted trees and pairwise distances revealed internal clustering rates of 43–79% across the dicots, but only 14–33% with monocot or gymnosperm ESTs. The frequencies of internal clustering with *Oryza* rooted trees overlapped both the dicot and gymnosperm values (which did not overlap one another), but the larger number of pairwise distances differentiated *Oryza* from the dicots.

The  $\gamma$  event appears to pre-date monocot–dicot divergence. Predominantly internal clustering (47–87%) was found for sequences from all angiosperms studied. For gymnosperms, internal clustering was less frequent (31–47%) but not significantly different from that for angiosperms. Relating the  $\gamma$  event to the angiosperm–gymnosperm divergence ~300 Myr ago<sup>23</sup> awaits more data.

Different models for *Arabidopsis* karyotypic evolution can be explored by considering the antiquity of duplication events, and the size of inferred duplicate chromosomal segments. The 'footprints' of the  $\alpha$  event of  $\leq 86$  Myr ago include 57 adjacent syntenic regions with opposite orientation and order explicable by localized inversions (Fig. 1 legend) that cover 89% of the genome, with an average length of about 9 centimorgans (cM). Using an estimated rate of structural mutations per chromosome pair per million years of divergence<sup>24</sup> ( $0.14 \pm 0.06$ ), and assuming that the present  $n = 5$  chromosomes has typified the *Arabidopsis* lineage in the past, about 60 rearrangements would be predicted, yielding modern chromosomal segments that average 8.3 cM in length. The possibility that  $n = 8$  better represents the taxon's history<sup>25</sup> would suggest greater divergence, with modern segments averaging 5.2 cM. If the lineage has varied between five and eight chromosomes, the fit of observed to predicted<sup>24</sup> values improves.

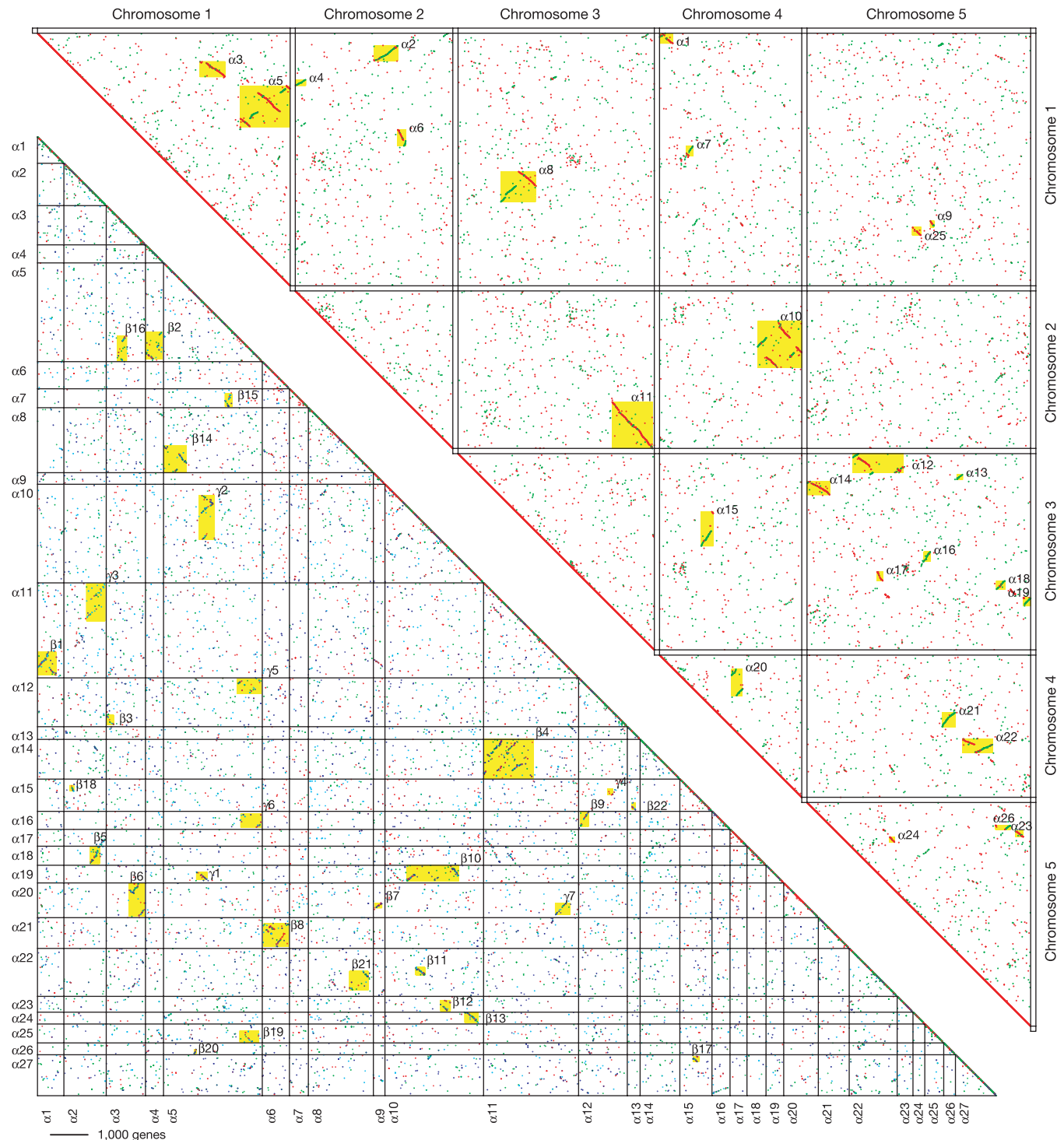
Establishing the provenance of ancient genome-wide duplication events revises and extends understanding of angiosperm evolutionary history, showing that much if not all of the flowering plant lineage is palaeo-polyploid. Although our  $\alpha$  event falls squarely in the range (65–100 Myr ago) during which much evidence supports a genome-wide duplication<sup>3,13–17</sup>, only two studies infer additional duplications. (1) Vision *et al.*<sup>15</sup> inferred four 'age classes' of duplications that bound 48, 39, 11 and 3% of the *Arabidopsis* genes, respectively, dismissing two additional classes as probable artefacts. However, <25% of all genes fall in two or more blocks, thus only one age class (C, 48%) could be inferred to involve most of the genome. Our  $\alpha$  event involves 89% of *Arabidopsis* genes, close to the total of all four age classes<sup>15</sup>. By inferring pre-duplication gene orders, then searching for more ancient duplication, we show that the  $\beta$  event involved  $\geq 51.6\%$  of the genes, more than the largest age class<sup>15</sup> and virtually all of which were also involved in the  $\alpha$  event. The  $\gamma$  event, which may be 100 Myr older than the most ancient age class<sup>15</sup>, covers more of the transcriptome (20.3%) than two of the



four age classes. (2) A recent study<sup>17</sup> using synonymous (third-nucleotide,  $K_s$ ) rather than overall<sup>15</sup> nucleotide substitution rates also contraindicates the four age classes, instead supporting our inference of two to three genome-wide duplications based on  $K_s$

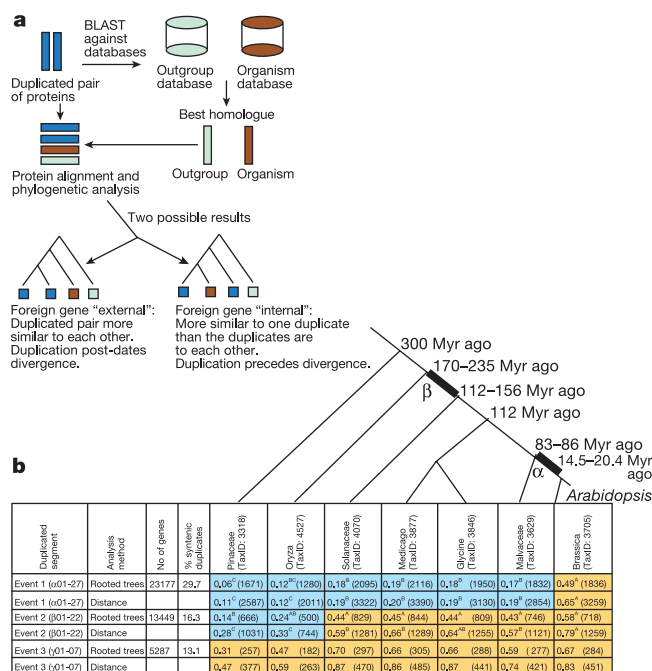
values similar to those found among our syntenic gene pairs ( $1.18 \pm 0.69$ ,  $2.33 \pm 1.03$  and  $2.82 \pm 1.16$  for  $\alpha$ ,  $\beta$  and  $\gamma$ , respectively, calculated as described<sup>26</sup>).

Using consensus among many gene trees to relate *Arabidopsis*



**Figure 1** Arrangement of duplicated protein-encoding genes in *Arabidopsis thaliana*. Top right,  $\alpha$  duplications. Both  $x$  and  $y$  axes represent 26,028 genes in their chromosomal order. The best-matching gene pairs are plotted, colour-coded to indicate same (red) or opposite (green) transcriptional orientations. For further analysis, 57 adjacent duplicated regions with opposite orientation and order explicable by localized inversions were combined into 26 'large' duplications ( $\alpha 01$ – $\alpha 26$ ) that each included  $\geq 1\%$  (260) of the genes. Eight shorter duplications were pooled ( $\alpha 27$ ). Lower left,  $\beta$  and  $\gamma$  duplications. Both  $x$  and  $y$  axes represent 21,749 genes, in an inferred ancestral order that accounts for

the composition of the 26 large  $\alpha$  duplications (at left and bottom). Twenty-nine  $\beta$  or  $\gamma$  duplications (see text) are highlighted. Colours show how the four modern *Arabidopsis* chromosome segments contribute to  $\beta$  or  $\gamma$  duplications, distinguishing contributions to the segments at left and bottom respectively from the: (1) lower-numbered chromosomes (red); (2) higher- and lower-numbered chromosomes (light blue); (3) lower- and higher-numbered chromosomes (dark blue); (4) higher-numbered chromosomes (green). Higher-resolution versions of the figure and lists of gene orders are available (see Supplementary Information).

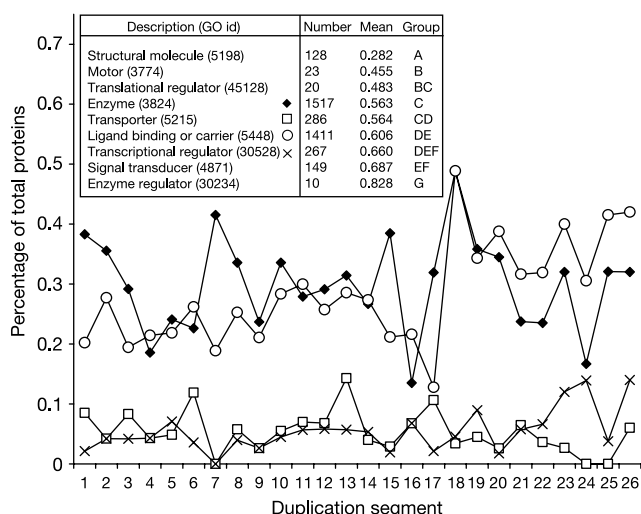


**Figure 2** Phylogenetic methodology for dating *Arabidopsis* duplications. **a**, Schematic of data flow for rooted gene tree and/or PAM-based pairwise protein distance analysis described in Methods. **b**, For each taxon (as listed) and analysis method (rooted trees, pairwise distances), the fraction of internal trees with superscripted letters indicating differences significant at  $P \geq 0.05$ , and number of trees that could be analysed (parentheses) are shown, and aligned with phylogenetic relationships among the taxa (above). Taxonomic divergence dates are as cited in the text. Detailed data for each individual  $\alpha$ ,  $\beta$  and  $\gamma$  segment pair are available (see Supplementary Information).

duplication events to its divergence from other angiosperms addresses significant pitfalls of 'dating' approaches based on molecular clocks<sup>15–17</sup>. Within each duplication event, much variation exists among syntenic gene pairs in the frequencies of internal/external clustering, but we question the interpretation<sup>15–17</sup> that such variation largely reflects the antiquity of duplication events. Individual chromosome segment pairs vary widely in the abundance of different classes of genes, which in turn have different degrees of conservation (Fig. 3). For example, structural proteins are well conserved, and enzyme regulators are especially diverse. Substantial differences in the extent of divergence between segment pairs (for example, average  $K_s$  values) can be explained simply by this sampling variation. A recent report<sup>21</sup> of  $K_s$  variation among plant genes of up to 14-fold, generally higher than previously reported, urges further caution in equating  $K_s$  with time.

A consensus of data from many genes is needed to 'date' duplication events, in that occasional false trees and/or distance estimates may occur due to rapid evolution of one duplicated copy, ancient proximal duplication followed by deletion of the true orthologue, failure to sample true orthologues in EST data, inter-locus convergence such as gene conversion, or other factors. For example, the Pinaceae–*Arabidopsis* divergence<sup>23</sup> far pre-dates the  $\alpha$  event, yet pine sequences show 6–11% internal trees.

By merging a phylogenetic approach to relating chromosomal duplications to the tree of life, with a genomic approach to mitigating information lost to diploidization, increased levels of angiosperm synteny (and associated opportunities for comparative biology) may be revealed. Virtually all previous comparative studies, including ours<sup>13</sup>, may have underestimated synteny between taxa by using 'one-to-one' comparisons to *Arabidopsis*<sup>5–11</sup>, which are appropriate only for closely related taxa that diverged from *Arabidopsis* after the  $\alpha$  event (such as other Brassicaceae).

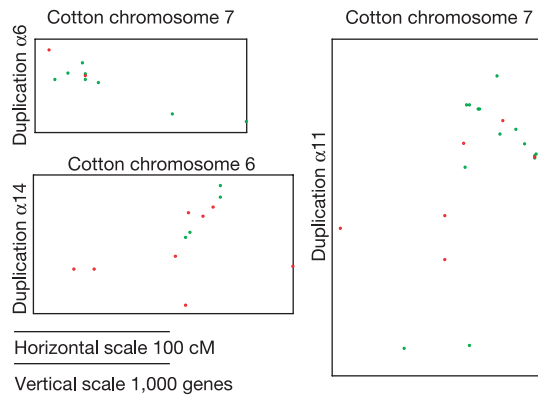


**Figure 3** Gene family composition contributes to different levels of divergence between duplicated chromosomal segments. Inset, Gene-Ontology-based molecular function annotations of syntenic duplicated gene pairs, including the number in each category, and mean protein similarity values (calculated by protidist). Tukey's analysis of square-root-transformed protein distances shows significant differences (A–G) in the degree of similarity for different gene families. Main panel, for  $\alpha$ 01– $\alpha$ 26 duplications, percentages of the total proteins (genes) are plotted for the five most abundant protein families. A duplicate pair can be in more than one family if different domains perform different functions, so the sum can exceed 100%.

Phylogenetic data provide a framework for making more appropriate intergenomic comparisons, by determining whether chromosomal duplications within taxa pre-date or post-date divergence among taxa. Although abundant genomic duplication makes this approach especially important in the angiosperms, it may contribute to better genomic comparisons in many lineages.

In large angiosperm genomes that are not likely to be sequenced soon (such as many major crops), much benefit might be gained from the sequences of botanical models by alignment of detailed genetic maps based upon conserved sequence-tagged sites. Comparison to a detailed cotton map (Fig. 4) showed that the two different members of most *Arabidopsis*  $\alpha$  segment pairs provide complementary synteny information. For each member of the 26 large  $\alpha$  segment pairs, we determined the number of 'best matches' (using BLASTN with an expectation value of  $E < 10^{-9}$ ) with genes on each of the 13 cotton chromosomes. To normalize over the different gene numbers on each cotton chromosome and *Arabidopsis* segment, we subtracted the random expectation (calculated by standard contingency methods). The paired *Arabidopsis* segments showed highly significant correlation ( $r = 0.197$ , 336 d.f.,  $P < 0.01$ ) of the residual frequencies of 'best matches' to the 13 cotton chromosomes. Comparisons of non-paired segments to one another showed no correlation ( $r = -0.02$ ). Only after assembling the two members of  $\alpha$  segment pairs into inferred ancestral orders did patterns of *Arabidopsis*–cotton synteny become clear (to be fully described elsewhere but see example, Fig. 4). In addition to its applied value for cotton improvement, this finding further supports the view that the  $\alpha$  event post-dated *Arabidopsis*–Malvaceae divergence. Using entire cotton chromosomes to assess the predictive value associated with pairing of duplicated *Arabidopsis* segments is very conservative, but is at present necessary. Sufficient data to analyse synteny in the 8.3-cM-average bins predicted<sup>24</sup> to persist after  $\leq 86$  Myr of divergence awaits mapping of more cotton genes.

Of special interest is synteny between monocots and eudicots, the two main branches of the angiosperms. One-to-one comparisons<sup>27</sup> show discernible parallels of the partial *Oryza* (rice) sequence to



**Figure 4** Examples of *Arabidopsis*–cotton synteny. Axes reflect relative gene orders (*Arabidopsis*) and map distances in cM (cotton). Cotton homoeologous series 7 corresponds in non-overlapping regions to *Arabidopsis* α06 (derived from parts of chromosomes 1 (red) and 2 (green)) and α11 (from chromosomes 2 (red) and 3 (green)) based on 10 and 20 matches respectively, numbers of matches expected to occur by chance at likelihoods of  $2.6 \times 10^{-5}$  and  $1.9 \times 10^{-3}$  based on  $\chi^2$  analysis. A sub-central region of cotton chromosome 6 corresponds to *Arabidopsis* α14 (from chromosomes 3, 5) based on 12 matches, and a chance likelihood of  $2.5 \times 10^{-5}$ .

*Arabidopsis*, although “... conservation is less extensive than previously predicted...” (in ref. 24). Factoring in two *Arabidopsis* duplications (α, β) plus at least one *Oryza* duplication<sup>27</sup> that post-date divergence of these taxa is likely to increase the level of conservation found.

By this synergistic approach, analysis and interpretation of whole-genome sequences benefits from partial sequence data for larger-genome taxa, in turn building contextual information important to using comparative approaches to accelerate analysis of larger genomes. Improved delineation of comparative gene arrangements promises new insights into fundamental questions, and invigorated progress towards practical applications of genomics. For example, the notion that particular gene arrangements may confer increased fitness has long awaited sufficient data to discern conserved organization of chromosomal segments that are larger, or contain more genes, than would be explicable by chance. Detailed study of the levels and patterns of diploidization and divergence in different gene families, and divergent taxa, may begin to shed light on the long-debated adaptive significance, if any, of polyploidy. Much knowledge exists about genomic changes that follow polyploidization<sup>12</sup>, but little is known about which specific events underlie the co-evolution of divergent genomes in a common nucleus to yield new phenotypes. Only 30% of *Arabidopsis* genes have retained syntenic copies in the ≤86 Myr since the α duplication—comparison to the 70% synteny of human and mouse proteins after 100 Myr (ref. 1) highlights the potential impact of gene loss on angiosperm evolution.

Better delineation of comparative gene arrangements may also aid in annotation of genomic sequences for other angiosperms, and also promises increased use of rapidly expanding knowledge about the action(s) of individual genes in facile models to identify convergent or parallel evolution of allelic variants important to agriculture or development. Such a comparative approach may especially accelerate progress towards identifying those genes that underlie complex physiological, morphological or behavioural phenotypes, and are discernible only as quantitative trait loci<sup>28</sup>. □

## Methods

### Duplication analysis

A total of 26,028 *Arabidopsis* gene sequences were downloaded from NCBI (<http://www.ncbi.nlm.nih.gov/>), encoded by their chromosomal order and transcriptional orientation, and compared to each other using BLASTP<sup>29</sup>. Only the top five, or two,

non-self protein matches that met a threshold of  $E < 10^{-10}$  were considered in α-duplication, or β/γ-duplication analysis, respectively. Circumscription of individual duplicated segments was largely as described<sup>13</sup>, using additional data to extend the limits of a duplication if a BLAST match revealed a pair of genes in consistent orientation within 20 or fewer genes (counted by combining both members of the pair) from the prior terminus. Adjacent syntenic regions with opposite orientation and order explicable by localized inversions were combined for further analyses, arbitrarily designating the linear order of the lower-numbered chromosomal region as ancestral, and assigning the inversion to the higher-numbered chromosome. In six cases, the proposed inversions involved two duplication regions.

### Gene tree analysis

Each duplicated syntenic gene pair was compared to each taxon-specific sequence (using nucleotide databases created by batch NCBI download of taxon IDs 3318, 3629, 3705, 3846, 3877, 4070 and 4527) using TBLASTN, results parsed, and the best match exceeding  $E < 10^{-5}$  translated to protein in the frame of the TBLASTN match, then (using BLAST) checked against all *Arabidopsis* proteins to confirm that the original genes were recovered. Rooted trees were made by including *Physcomitrella* (taxon ID 3217) sequences. The most similar regions between a duplicated pair, the best-matching homologue and the moss outgroup (where relevant) were determined using pairwise alignment as implemented by the ‘water’ program in EMBOSS (<http://www.emboss.org/>), then aligned using CLUSTALW version 1.82 and the default parameters (PAM matrix; gap opening penalty = 10.0; gap extension penalty = 0.2). For rooted analyses, 100 bootstrap replicates were created using the EMBOSS interface to the ‘seqboot’ program in PHYLIP version 3.6 (<http://evolution.genetics.washington.edu/phylip.html>), and protein parsimony trees created using the ‘protpars’ program in PHYLIP (default parameters). A consensus tree rooted with moss was created using the EMBOSS interfaces to the ‘consense’ program in PHYLIP, and the resulting tree file was parsed to determine the position of the taxon-specific sequence relative to the *Arabidopsis* duplicates. For non-rooted analyses, protein distances for *Arabidopsis* duplicates and the other taxon were calculated using ‘protdist’ (PAM matrix, default parameters) in PHYLIP, and compared to determine the most similar pair. NCBI downloads, BLAST parsing, sequence translation and phylogenetic analyses were automated using Python scripts available from <http://www.plantgenome.uga.edu/project-bioinformatics.htm>.

The fractions of ‘internal trees’ associated with each duplication cycle were compared using one-way analysis of variance (ANOVA) for correlated samples and Tukey’s HSD analysis for post-ANOVA comparisons between organisms (R. Lowry, chapter 15 in <http://faculty.vassar.edu/lowry/webtext.html>). Individual duplicated segment pairs were considered treatments, and the indicated taxa were conditions, accounting for correlations that may result from comparing identical genes in different taxa. This is conservative, because in many cases different regions of an *Arabidopsis* gene matched ESTs from different taxa, reducing the correlation problem.

### Arabidopsis–cotton synteny

The order of 2,102 sequence-tagged sites (GenBank accession numbers provided in Supplementary Information) along the chromosomes of a hypothetical cotton progenitor pre-dating A/D-subgenome divergence was inferred by using anchor probes mapped in both diploid and tetraploid cottons<sup>30</sup> plus new loci to interpolate the locations of sequence-tagged sites mapped in only a subset of these populations. This map was compared to the inferred ancestral gene order pre-dating the *Arabidopsis* α duplication by finding the most similar BLASTN match of  $E < 10^{-9}$  to any *Arabidopsis* gene.

Received 25 October 2002; accepted 5 February 2003; doi:10.1038/nature01521.

1. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
2. Smith, S. S. *et al.* Analyses of the extent of shared synteny and conserved gene orders between the genome of *Fugu rubripes* and Human 20q. *Genome Res.* **12**, 776–784 (2002).
3. The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
4. Yang, Y.-W., Lai, K.-N., Tai, P.-Y. & Li, W.-H. Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other angiosperm lineages. *J. Mol. Evol.* **48**, 597–604 (1999).
5. Salse, J., Benoit, P., Cooke, R. & Delseny, M. Synteny between *Arabidopsis thaliana* and rice at the genome level: a tool to identify conservation in the ongoing Rice Genome Sequencing Project. *Nucleic Acids Res.* **30**, 2317–2328 (2002).
6. Mayer, K. *et al.* Conservation of microstructure between a sequenced region of the genome of rice and multiple segments of the genome of *Arabidopsis thaliana*. *Genome Res.* **11**, 1167–1174 (2001).
7. Ku, H., Vision, T., Liu, J. & Tanksley, S. D. Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny. *Proc. Natl Acad. Sci. USA* **97**, 9121–9126 (2000).
8. Grant, D., Cregan, P. & Shoemaker, R. C. Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **97**, 4168–4173 (2000).
9. Lee, J. M., Grant, D., Vallejos, C. E. & Shoemaker, R. C. Genome organization in dicots. II. *Arabidopsis* as a ‘bridging species’ to resolve genome evolution events among legumes. *Theor. Appl. Genet.* **103**, 765–773 (2001).
10. Rossberg, M. *et al.* Comparative sequence analysis reveals extensive microcolinearity in the lateral suppressor regions of the tomato, *Arabidopsis*, and Capsella genomes. *Plant Cell* **13**, 979–988 (2001).
11. Liu, H., Sachidanandam, R. & Stein, L. Comparative genomics between rice and *Arabidopsis* shows scant collinearity in gene order. *Genome Res.* **11**, 2020–2026 (2001).
12. Eckhardt, N. A sense of self: The role of DNA sequence elimination in allopolyploidization. *Plant Cell* **13**, 1699–1704 (2001).
13. Paterson, A. H. *et al.* Comparative genomics of plant chromosomes. *Plant Cell* **12**, 1523–1539 (2000).



14. Blanc, G., Barakat, A., Guyot, R., Cooke, R. & Delseny, M. Extensive duplication and reshuffling in the *Arabidopsis* genome. *Plant Cell* **12**, 1093–1101 (2000).
15. Vision, T. D., Brown, D. B. & Tanksley, S. D. The origins of genomic duplications in *Arabidopsis*. *Science* **290**, 2114–2117 (2000).
16. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155 (2000).
17. Simillion, C., Vandepoele, K., Van Montagu, M. C. E., Zabeau, M. & Van de Peer, Y. The hidden duplication past of *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **99**, 13627–13632 (2002).
18. Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R. & Polasky, S. Agricultural sustainability and intensive production practices. *Nature* **418**, 671–677 (2002).
19. Bennett, M. D. & Smith, J. B. Nuclear DNA amounts in angiosperms. *Proc. R. Soc. Lond. B* **274**, 227–274 (1976).
20. Strauss, E. Can mitochondrial clocks keep time? *Science* **283**, 1435–1438 (1999).
21. Zhang, L., Vision, T. J. & Gaut, B. S. Patterns of nucleotide substitution among simultaneously duplicated gene pairs in *Arabidopsis thaliana*. *Mol. Biol. Evol.* **19**, 1464–1473 (2002).
22. Benton, M. J. *The Fossil Record 2* (Chapman and Hall, New York, 1993).
23. Bowe, L. M., Coat, G. & dePamphilis, C. W. Phylogeny of seed plants based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proc. Natl Acad. Sci. USA* **97**, 4092–4097 (2000).
24. Paterson, A. H. et al. Toward a unified genetic map of higher plants, transcending the monocot–dicot divergence. *Nature Genet.* **14**, 380–382 (1996).
25. Koch, M., Bishop, J. & Mitchell-Olds, T. Molecular systematics and evolution of *Arabidopsis* and *Arabis*. *Plant Biol.* **1**, 529–537 (1999).
26. Yang, Z. & Nielsen, R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17**, 32–43 (2000).
27. Goff, S. A. et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**, 92–100 (2002).
28. Paterson, A. H. et al. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* **335**, 721–726 (1988).
29. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
30. Brubaker, C. L., Paterson, A. H. & Wendel, J. F. Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome* **42**, 184–203 (1999).

**Supplementary Information** accompanies the paper on Nature's website  
(<http://www.nature.com/nature>).

**Acknowledgements** We thank A. Feltus, J. C. Kissinger and S. Schulze for comments on the manuscript, and the Paterson Lab for technical support. This work was supported by the US Department of Agriculture National Research Initiative and Initiative for Future Agriculture and Food Safety, the US National Science Foundation Plant Genome Research Program, the Howard Hughes Medical Institute Graduate Fellowship Program, the International Consortium for Sugarcane Biotechnology, the Georgia Cotton Commission/Cotton Inc., and the Georgia Agricultural Experiment Station.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to A.H.P.  
(e-mail: paterson@uga.edu).

## The role of presenilin cofactors in the $\gamma$ -secretase complex

**Nobumasa Takasugi<sup>\*</sup>, Taisuke Tomita<sup>\*</sup>, Ikuo Hayashi<sup>\*</sup>,  
Makiko Tsuruoka<sup>\*</sup>, Manabu Niimura<sup>\*</sup>, Yasuko Takahashi<sup>\*</sup>,  
Gopal Thinakaran<sup>†</sup> & Takeshi Iwatsubo<sup>\*</sup>**

<sup>\*</sup> Department of Neuropathology and Neuroscience, Graduate School of Pharmaceutical Sciences, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

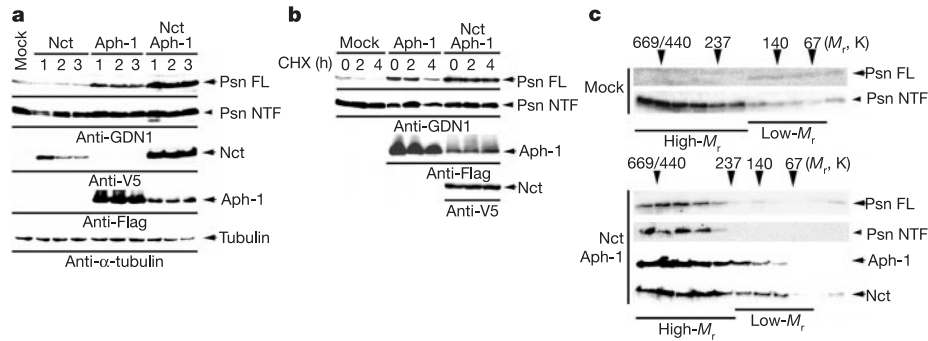
<sup>†</sup> Department of Neurobiology, Pharmacology and Physiology, The University of Chicago, 924 East 57th Street, Chicago, Illinois 60637, USA

Mutations in presenilin genes account for the majority of the cases of the familial form of Alzheimer's disease (FAD). Presenilin is essential for  $\gamma$ -secretase activity, a proteolytic activity involved in intramembrane cleavage of Notch and  $\beta$ -amyloid precursor protein ( $\beta$ APP)<sup>1,2</sup>. Cleavage of  $\beta$ APP by FAD mutant presenilin results in the overproduction of highly amyloidogenic amyloid  $\beta$ 42 peptides<sup>3–6</sup>.  $\gamma$ -Secretase activity requires the formation of a stable, high-molecular-mass protein complex<sup>7–11</sup> that,

in addition to the endoproteolysed fragmented form of presenilin, contains essential cofactors including nicastrin<sup>12–14</sup>, APH-1 (refs 15–18) and PEN-2 (refs 16, 19). However, the role of each protein in complex formation and the generation of enzymatic activity is unclear. Here we show that *Drosophila* APH-1 (Aph-1) increases the stability of *Drosophila* presenilin (Psn) holoprotein in the complex. Depletion of PEN-2 by RNA interference prevents endoproteolysis of presenilin and promotes stabilization of the holoprotein in both *Drosophila* and mammalian cells, including primary neurons. Co-expression of *Drosophila* Pen-2 with Aph-1 and nicastrin increases the formation of Psn fragments as well as  $\gamma$ -secretase activity. Thus, APH-1 stabilizes the presenilin holoprotein in the complex, whereas PEN-2 is required for endoproteolytic processing of presenilin and conferring  $\gamma$ -secretase activity to the complex.

Presenilin is essential for  $\gamma$ -secretase cleavage, which releases amyloid  $\beta$ -peptide (A $\beta$ ) and the intracellular domain of Notch by intramembraneous proteolysis of  $\beta$ -amyloid precursor protein ( $\beta$ APP) and Notch, respectively<sup>1,2</sup>. Presenilin mediates  $\gamma$ -secretase function by forming a highly stable protein complex of high relative molecular mass (high- $M_r$ ) together with a set of cofactor proteins<sup>7–11</sup>. In addition to nicastrin (NCT), a type I single-pass membrane glycoprotein<sup>12</sup>, two additional putative presenilin cofactors have been identified: APH-1, a multi-transmembrane protein coded by a gene whose deletion leads to hypoplasia of the anterior pharynx in *Caenorhabditis elegans*, was found to be a Notch pathway member possibly involved in presenilin function<sup>15</sup>; aph-1 was also identified as one of the presenilin enhancer genes (*pen-1*) together with *pen-2*, which codes for a double-membrane-spanning protein<sup>16</sup>. NCT, APH-1 and PEN-2 are required for  $\gamma$ -secretase function and accumulation of presenilin fragments<sup>12–14,16–19</sup>, although the differential roles of each cofactor in the formation of the high- $M_r$  presenilin protein complex, and whether NCT, APH-1 and PEN-2 represent the principal presenilin cofactors to confer  $\gamma$ -secretase activity, remain elusive.

We stably transfected *Drosophila* S2 cells with Aph-1, and found a significant increase in the levels of endogenous Psn holoprotein. This was markedly enhanced by expression of Aph-1 and *Drosophila* NCT (Nct), but was not observed in cells transfected with Nct alone. The levels of Psn fragments involved in the active form of  $\gamma$ -secretase were not altered in either case (Fig. 1a). We next treated stably co-transfected S2 cells overexpressing Aph-1 and Nct with cycloheximide (CHX) to block total cellular protein synthesis, and examined the stability of Psn and other proteins. In mock-transfected S2 cells (transfected with an empty vector alone) only small amounts of Psn holoprotein were detectable, which rapidly degraded within about 4 h of CHX treatment; however, fragments of Psn were relatively more abundant and highly stable, in a manner similar to that of mammalian presenilin<sup>7,11</sup> (Fig. 1b, left panel). In contrast, Psn holoprotein levels were significantly increased by overexpression of Aph-1 or of Aph-1 and Nct, and remained highly stabilized (Fig. 1b, middle and right panels, respectively). Levels of Aph-1 and Nct in co-transfected S2 cells were also highly stable during the period of CHX treatment, suggesting that Psn (including holoprotein) forms a highly stabilized protein complex together with Aph-1 and Nct under these conditions (Fig. 1b, middle and right panels). To gain support for this hypothesis, we then separated CHAPSO-solubilized membrane fractions of S2 cells stably transfected with Aph-1 and Nct by glycerol velocity gradient centrifugation<sup>10,11</sup>. Psn holoprotein in cells overexpressing Aph-1 and Nct was fractionated totally in high- $M_r$  ranges of 232–440K together with Psn fragments (Fig. 1c, lower panel, FL). This was in contrast to the exclusive low- $M_r$  distribution of short-lived Psn holoproteins in S2 cells in the absence of Aph-1 overexpression (Fig. 1c, upper panel, FL). Furthermore, most of the Aph-1 and Nct proteins were found in the high- $M_r$  fractions (Fig. 1c, lower panel). Taken together, our data support the hypothesis that Aph-1 represents



**Figure 1** Effects of Aph-1 on accumulation, stabilization and high- $M_r$  complex formation of Psn holoprotein. **a**, Immunoblot analysis of S2 cells stably expressing Nct, Aph-1 or both Nct and Aph-1 to determine expression of Psn N-terminal fragment (NTF), Nct and Aph-1. Three independent stable cell lines for each cDNA are shown. Note that Nct appears as almost a single band, unlike mammalian NCT, and that the levels of

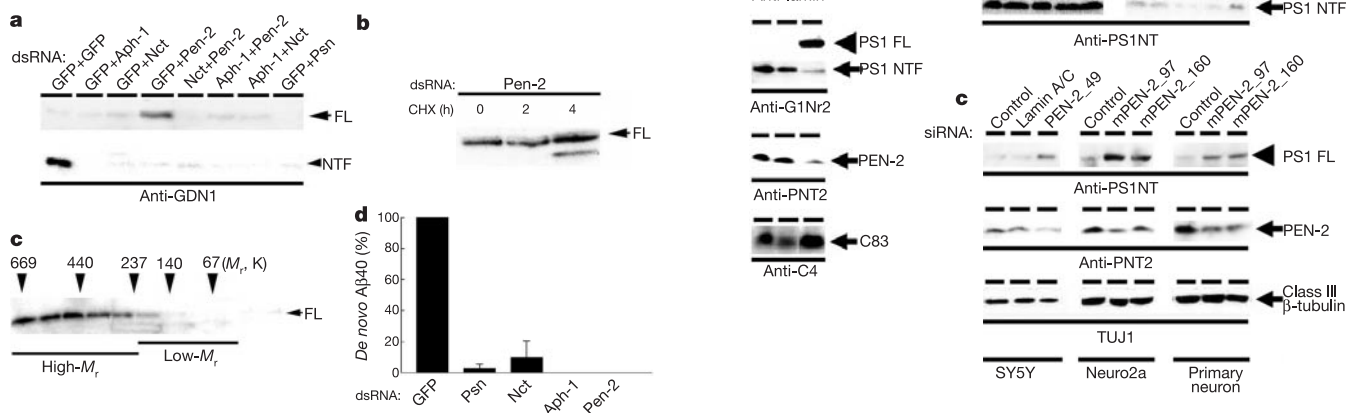
Aph-1 are downregulated on co-expression of Nct. **b**, Analysis of the stability of Psn, Nct or Aph-1 on treatment with cycloheximide (CHX) and chase for 2–4 h in S2 cells stably expressing Aph-1 alone (middle lanes) or Nct and Aph-1 (right lanes). **c**, Glycerol velocity separation of Psn holoprotein (FL) or NTF, Aph-1 and Nct in stable S2 cells.

the major ‘stabilizing’ cofactor of Psn that promotes the key step of stable high- $M_r$  complex formation. However, the level of A $\beta$  secretion from cells with or without co-expression of Nct and Aph-1 was not significantly different, suggesting that some additional factors are needed for the upregulation of  $\gamma$ -secretase activity.

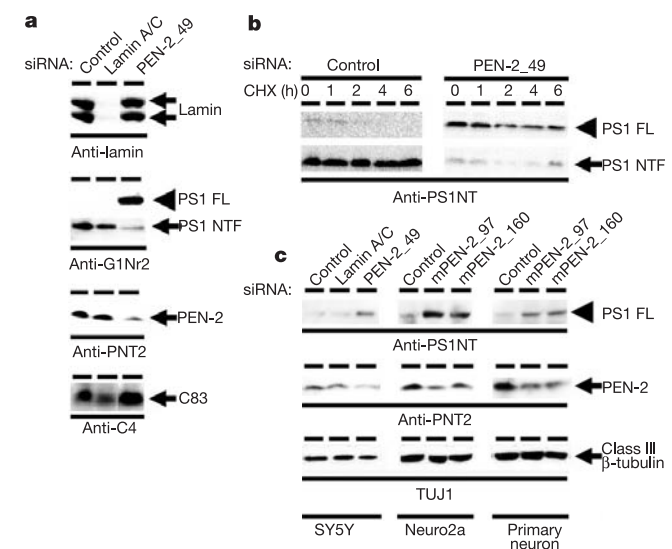
To examine the role of PEN-2—another putative cofactor of presenilin, which harbours two membrane-spanning domains—in the formation and function of the  $\gamma$ -secretase complex, we performed double-stranded RNA-mediated interference (RNAi) experiments. Unlike the results from Aph-1 or Nct RNAi<sup>13,16</sup>, we found that treatment of S2 cells with Pen-2 RNAi leads to loss of fragment forms of Psn and accumulation of Psn holoprotein (Fig. 2a; see also Supplementary Fig. S1). Application of CHX to S2 cells treated with Pen-2 RNAi showed that the accumulated Psn holoprotein is highly stabilized (Fig. 2b) and is recoverable in high- $M_r$  fractions (Fig. 2c). In contrast to the result of the Pen-2 RNAi experiment, combinations of Pen-2 and Aph-1 or Pen-2 and Nct RNAi abolished the accumulation of either fragment or holoprotein forms of Psn (Fig. 2a).  $\gamma$ -Secretase activities, as determined by an *in vitro* A $\beta$  generation assay using recombinant  $\beta$ APP

C100 (carboxy-terminal 99 amino acid residues of human  $\beta$ APP) tagged with Flag, Myc and His as a substrate, were totally abolished by every combination of RNAi that diminished the Psn fragment (Fig. 2d). These data provide genetic evidence to support the hypothesis that Nct and Aph-1 act upstream of Pen-2 in the formation of the  $\gamma$ -secretase complex. The accumulation of Psn holoprotein by Pen-2 RNAi was also observed in BG2 cells, a *Drosophila* neuronal cell line derived from primary cultures of larval central nervous system<sup>20</sup> (Supplementary Fig. S2).

To examine further whether PEN-2 has a similar role in the formation of the  $\gamma$ -secretase complex in mammalian cells, we performed PEN-2 RNAi by transfection of small interfering RNA (siRNA)<sup>21</sup>. Knockdown of PEN-2 expression in human HeLa cells resulted in the accumulation of highly stabilized presenilin 1 (PS1, also known as PSEN1) holoproteins accompanied by a decrease in



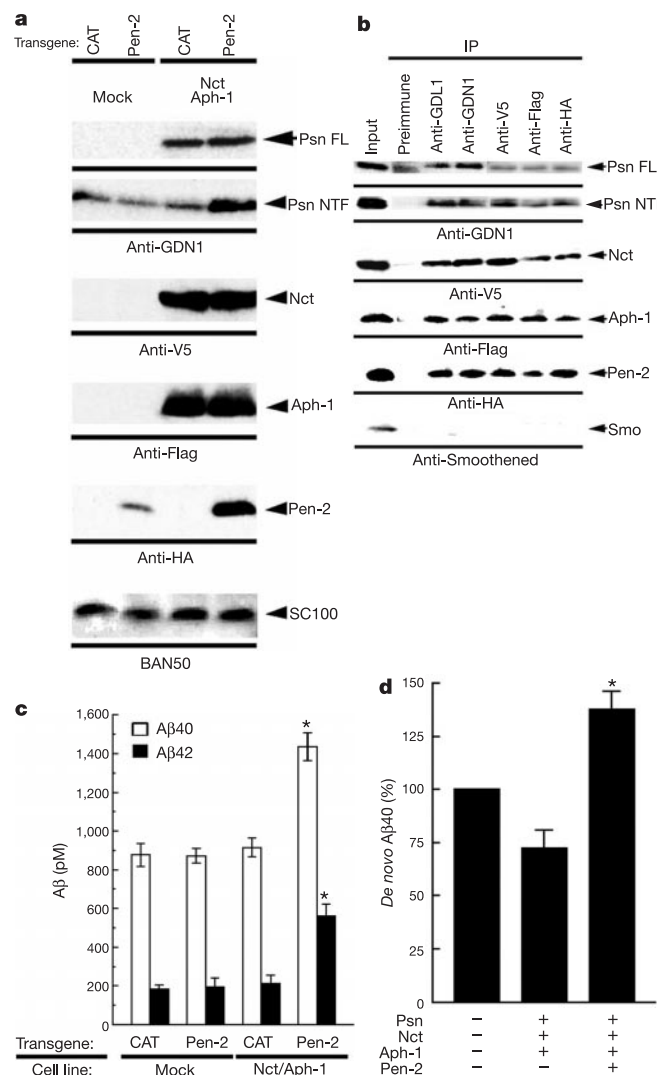
**Figure 2** Role of Pen-2 in the metabolism of Psn and  $\gamma$ -secretase activity. **a**, RNAi inactivation of Psn cofactors. S2 cells were treated by combinations of two dsRNAs, and the levels of Psn FL or NTF were analysed by immunoblotting. **b**, Increased stability of Psn holoprotein by Pen-2 RNAi as shown by CHX treatment. **c**, Glycerol velocity separation of Psn holoprotein from Pen-2 RNAi-treated cells. **d**, Relative levels of A $\beta$ 1–40 peptides (mean  $\pm$  s.e.,  $n = 3$ ) generated from recombinant C100 by *in vitro* co-incubation with membrane fractions of S2 cells treated by RNAi for green fluorescent protein (GFP) (= 100%), Psn, Nct, Aph-1 or Pen-2.



**Figure 3** Role of PEN-2 in the metabolism of mammalian PS1. **a**, siRNA-mediated RNAi of human PEN-2 in HeLa cells. HeLa cells were transfected with siRNA duplexes (indicated above each lane) twice during 6 days of culture and lysates were analysed by immunoblotting with antibodies shown below each panel. **b**, Increased stability of PS1 holoprotein by PEN-2 RNAi as shown by CHX treatment. **c**, PEN-2 RNAi in neuronal cells. SY5Y, Neuro2a and mouse primary neurons (DIV2) were treated with siRNA for 3 days, and the levels of PS1 holoproteins and of PEN-2 proteins were analysed by immunoblotting. The expression of neuronal class III  $\beta$ -tubulin was monitored by an antibody TUJ1 as a control protein in neuronal cells. mPEN-2, mouse PEN-2.

the levels of PS1 fragments as well as of  $\gamma$ -secretase activity, as evidenced by the accumulation of  $\beta$ APP C83 (Fig. 3a, b). We confirmed that RNAi depletion of PEN-2 caused an accumulation of PS1 holoproteins in mammalian neuronal cells (that is, human SH-SY5Y and mouse neuro2a neuroblastoma cells as well as mouse primary cultured cortical neurons) (Fig. 3c). Overexpression of human PEN-2 in Pen-2 RNAi-treated *Drosophila* S2 cells rescued the diminution in Psn fragments and attenuated the accumulation of Psn holoproteins (Supplementary Fig. S3).

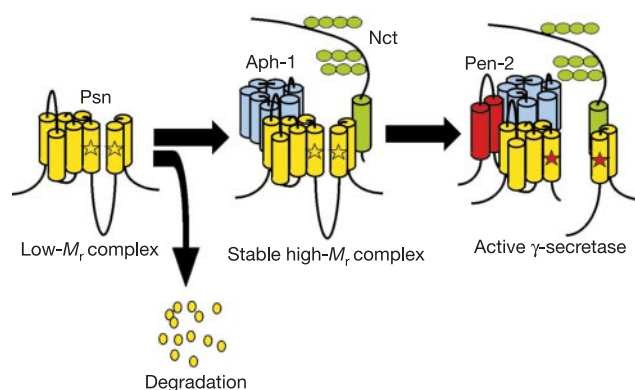
We next performed transient overexpression of Pen-2 in S2 cells stably transfected with Aph-1 and Nct, and observed an increase in



**Figure 4** Overexpression of Pen-2 in addition to Nct and Aph-1 increases the accumulation of Psn fragments and the  $\gamma$ -secretase activity. **a**, S2 cells stably expressing Nct and Aph-1 (right two lanes) or an empty vector (left two lanes) were transiently transfected with chloramphenicol acetyltransferase (CAT) or Pen-2, together with SC100, and analysed by immunoblotting. **b**, Co-immunoprecipitation (IP) of Psn, Nct, Aph-1 and Pen-2. Note that cofactor proteins specifically bind to Psn but not to Smoothed (Smo), an abundantly expressed endogenous membrane protein. **c**, The levels of secreted A $\beta$ 40 and A $\beta$ 42 peptides (mean  $\pm$  s.e.,  $n = 3$ ) from S2 cells transfected with combinations of Nct, Aph-1 or Pen-2 and SC100 as in **a** were quantified by A $\beta$  ELISA. Asterisk, statistically significant by analysis of variance (ANOVA;  $P < 0.01$ ). **d**, Relative levels of A $\beta$ 1–40 peptides (mean  $\pm$  s.e.,  $n = 6$ ; A $\beta$  level in mock-transfected cells is 100%) generated from recombinant C100 by *in vitro* co-incubation with membrane fractions of S2 cells transfected with combinations of Psn, Nct, Aph-1 or Pen-2 (see Supplementary Fig. S5 for the overproduction of Psn fragments on co-transfection of Pen-2 in this experiment). Asterisk, statistically significant by ANOVA ( $P < 0.005$ ).

the accumulation of Psn fragments (Fig. 4a, upper panel, Psn NTF). To determine whether Pen-2, Aph-1, Nct and Psn function by directly binding to each other, we performed co-immunoprecipitation studies in CHAPSO-solubilized membrane fractions of S2 cells stably co-transfected with the three cofactors, and found that any two of the components among Psn (holoprotein as well as fragments), Nct, Aph-1 or Pen-2 are co-immunoprecipitated (Fig. 4b). In addition, Pen-2 was fractionated in the high- $M_r$  ranges together with Psn, Nct or Aph-1 by glycerol velocity gradient centrifugation (Supplementary Fig. S4), suggesting that they function together and form a protein complex. Finally, we found that secretion of A $\beta$  from S2 cells (Fig. 4c), as well as the total cellular  $\gamma$ -secretase activity as determined by an *in vitro* A $\beta$  generation assay using recombinant  $\beta$ APP C100 (Fig. 4d; see also Supplementary Fig. S5) was significantly increased by the co-transfection of Pen-2, indicating that  $\gamma$ -secretase activity was augmented by triple co-expression of Aph-1, Nct and Pen-2. Taken together, our data indicate that Pen-2 is a cofactor that is required for the final step of presenilin complex maturation after presenilin is stabilized in a holoprotein form and incorporated into a high- $M_r$  protein complex, conferring  $\gamma$ -secretase activities and leading to endoproteolysis of presenilin.

By combining RNAi inactivation and stable overexpression of cofactor proteins of Psn in *Drosophila* S2 cells, we have been able to dissect the process of high- $M_r$  protein complex formation and the activation of  $\gamma$ -secretase function of presenilin (Fig. 5). Overexpression of Aph-1 was sufficient to elicit the stabilization and high- $M_r$  complex formation of endogenous presenilin in a holoprotein form, although this high- $M_r$  protein complex harbouring presenilin holoprotein is still inactive with respect to  $\gamma$ -secretase (Fig. 5, middle panel). The final maturation step of the presenilin complex, which consists of presenilin fragments and demonstrates  $\gamma$ -secretase activity, may then be mediated by the additional function of Pen-2 through directly interacting with the Psn–Aph-1–Nct complex<sup>14</sup> (Fig. 5, right panel). This model is supported by the following observations: (1) Pen-2 RNAi facilitated the accumulation of stabilized Psn holoprotein, as is seen with Aph-1 overexpression; (2) Psn polypeptide accumulation is totally abolished by RNAi-mediated inactivation of Nct or Aph-1 expression; and (3) co-expression of Pen-2, Nct and Aph-1 rendered the stabilized Psn holoprotein into the  $\gamma$ -secretase-active fragment form. The latter result also suggests that APH-1, NCT and PEN-2 represent the set of major presenilin cofactors sufficient for the formation of the



**Figure 5** Schematic depiction of the stepwise assembly and activation of the  $\gamma$ -secretase complex in *Drosophila* cells. Nascent Psn holoprotein forms a low- $M_r$  protein complex and is rapidly degraded, whereas a fraction of Psn is stabilized by binding to Aph-1 (and Nct), although it still remains  $\gamma$ -secretase inactive. Pen-2 elicits the final step of maturation of the  $\gamma$ -secretase complex, facilitating endoproteolysis of Psn and conferring  $\gamma$ -secretase activity. Coloured tubes represent the putative transmembrane domains (TMD) of each protein, and stars within the sixth and seventh TMD of presenilin symbolize active (red) and inactive (unfilled) aspartate residues involved in  $\gamma$ -secretase activities.



$\gamma$ -secretase complex; a large protease complex reminiscent of the proteasome, which executes the cleavage of peptide bonds buried within the lipid bilayer<sup>22,23</sup>. Further efforts to reconstitute  $\gamma$ -secretase activity from recombinant proteins of the presenilin and the three cofactors *in vitro* will be needed to elucidate the configuration and function of presenilin complex and to develop therapeutic approaches to Alzheimer's disease based on inhibition of  $\gamma$ -secretase. □

## Methods

### Construction of expression plasmids

Complementary DNAs encoding Aph-1 or Nct were obtained from IMAGE clone (clone number LD12037 and LD05644, respectively; Invitrogen), and subcloned into a pAc5.1/V5-His A vector (Invitrogen) with Flag or V5 tags at the C terminus, respectively. Pen-2 or human PEN-2 cDNAs tagged with haemagglutinin (HA) at the amino terminal were amplified from a *Drosophila* or human brain cDNA libraries (Stratagene) and subcloned into a pIB/V5-His TOPO vector (Invitrogen). A cDNA encoding the C-terminal 99 amino acids of human  $\beta$ APP fused to a signal peptide of rat pre-proenkephalin at the N terminus (SC100)<sup>6</sup> was subcloned into pAc5.1/V5-His A vector. A cDNA encoding C100-Flag/Myc/His in pTrcHis2A vector (Invitrogen) was generated by polymerase chain reaction (PCR) using a cDNA encoding C100 fused to Flag at the C terminus as a template. All constructs were sequenced using Thermosequenase (Amersham) on an automated sequencer (Li-Cor).

### Cell culture and transfection

*Drosophila* Schneider (S2) or BG2 cells were maintained in Schneider's insect medium (Sigma) supplemented with 10% fetal bovine serum, 5% peptone and penicillin/streptomycin at 24 °C. Transient transfection of cDNAs into S2 or BG2 cells was performed using cellfectin (Invitrogen) according to the manufacturer's instructions, and cells or culture medium were collected after 48 h transfection. Stable transformants in S2 cells were generated by transfection of cDNAs in pAc5.1/V5-His A vector together with pCoHygro or pCoBlast vector (Invitrogen) (ratio of transfected cDNAs, 2:0.1  $\mu$ g) using cellfectin, and selected in S2 medium containing hygromycin (Wako) or blastcidin (Invitrogen) at 250 or 5  $\mu$ g ml<sup>-1</sup>, respectively. HeLa, SY5Y, Neuro2a cells<sup>8,10</sup> and mouse primary cortical neurons derived from E16 fetuses of BALB/c mice<sup>24</sup> were maintained as described.

### RNAi

For the generation of the double-stranded RNAs (dsRNA), transcription templates flanked by T7 RNA promoter sequences at each end were generated by PCR using oligonucleotides containing sequences for the T7 RNA polymerase-binding site. dsRNAs were prepared from transcription templates by using the MEGAscript T7 kit (Ambion) and transfected into S2 or BG2 cells using cellfectin as in transfection of cDNA plasmids. Small interfering RNA (siRNA) duplexes for human or mouse PEN-2 (nucleotide sequences shown in Supplementary Table) were purchased from JBIOS and transfected by GeneSilencer (Gene Therapy Systems) for HeLa, SY5Y and Neuro2a cells or oligofectamine (Invitrogen) for primary neurons<sup>21</sup>. Cell lysates and conditioned medium were collected after incubation for indicated times.

### Antibodies, immunoblot analysis and fractionation studies

Two rabbit polyclonal antibodies to Psn—anti-GDN1 against glutathione S-transferase (GST) fused to amino acids 2–52 of Psn and anti-GDL1 against GST fused to amino acids 358–426 of Psn—were generated<sup>11</sup>. Anti-G1Nr2 (ref. 10) and PS1NT<sup>25</sup> against GST fused to the N-terminal region of human PS1 have been described previously. A rabbit anti-PEN-2 antibody PNT2 was raised against a synthetic peptide corresponding to the N-terminal 26 amino acids of human/mouse PEN-2. A rabbit polyclonal antibody anti-C4 against the cytoplasmic C terminus of human  $\beta$ APP and a rat anti-smoothed antibody are from Y. Ihara and S. Cohen, respectively. Other antibodies were purchased from BAbCO (anti-neuronal class III  $\beta$ -tubulin, clone TUJ1), Chemicon (anti-lamin A/C, clone JoL2), Invitrogen (anti-V5), Roche (anti-HA) or Sigma (anti-Flag, anti- $\alpha$ -tubulin and anti-V5 polyclonal). Preparation of cell lysates and membrane fractions, immunoblot analysis, cycloheximide treatment, glycerol velocity centrifugation, immunoprecipitation or quantification of A $\beta$  by two-site enzyme-linked immunosorbent assays (ELISAs) were performed as previously described<sup>16,10,11,24,26–28</sup>.

### *In vitro* $\gamma$ -secretase assay

Recombinant C100-Flag/Myc/His protein was expressed in *Escherichia coli*, solubilized by 6 M guanidine hydrochloride in phosphate buffer (20 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.2, 500 mM NaCl) containing complete/EGTA-free protease inhibitor cocktail (Roche), and purified by nickel chelating affinity chromatography on Akta Prime (Amersham). All procedures were performed at 4 °C. Purified recombinant C100-Flag/Myc/His was stored at –80 °C until use. For *in vitro* assay of  $\gamma$ -secretase activities, C100-Flag/Myc/His (1  $\mu$ M) was incubated with the S2 cell membrane fraction (250  $\mu$ g ml<sup>-1</sup>) in  $\gamma$ -buffer (HEPES buffer containing 0.25% CHAPSO, 5 mM EDTA, 5 mM 1,10-phenanthroline, 10 mg ml<sup>-1</sup> phosphoramidon and complete protease inhibitor cocktail) at 4 or 37 °C for 3 h. The reaction was stopped by adding 10% BIGCHAP (final concentration of 1%) and boiling for 2 min. The samples were ultracentrifuged and the supernatants were analysed by two-site ELISAs for *in vitro* generation of A $\beta$ .

Received 30 December 2002; accepted 19 February 2003; doi:10.1038/nature01506.  
Published online 16 March 2003.

- De Strooper, B. *et al.* Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein. *Nature* **391**, 387–390 (1998).
- De Strooper, B. *et al.* A presenilin-1-dependent  $\gamma$ -secretase-like protease mediates release of Notch intracellular domain. *Nature* **398**, 518–522 (1999).
- Duff, K. *et al.* Increased amyloid- $\beta$ 42(43) in brains of mice expressing mutant presenilin 1. *Nature* **383**, 710–713 (1996).
- Borchelt, D. R. *et al.* Familial Alzheimer's disease-linked presenilin 1 variants elevate A $\beta$ 1-42/1-40 ratio *in vitro* and *in vivo*. *Neuron* **17**, 1005–1013 (1996).
- Citron, M. *et al.* Mutant presenilins of Alzheimer's disease increase production of 42-residue amyloid  $\beta$ -protein in both transfected cells and transgenic mice. *Nature Med.* **3**, 67–72 (1997).
- Tomita, T. *et al.* The presenilin 2 mutation (N141I) linked to familial Alzheimer disease (Volga German families) increases the secretion of amyloid  $\beta$  protein ending at the 42nd (or 43rd) residue. *Proc. Natl Acad. Sci. USA* **94**, 2025–2030 (1997).
- Ratovitski, T. *et al.* Endoproteolytic processing and stabilization of wild-type and mutant presenilin. *J. Biol. Chem.* **272**, 24536–24541 (1997).
- Thinakaran, G. *et al.* Evidence that levels of presenilins (PS1 and PS2) are coordinately regulated by competition for limiting cellular factors. *J. Biol. Chem.* **272**, 28415–28422 (1997).
- Capell, A. *et al.* The proteolytic fragments of the Alzheimer's disease-associated presenilin-1 form heterodimers and occur as a 100–150-kDa molecular mass complex. *J. Biol. Chem.* **273**, 3205–3211 (1998).
- Tomita, T. *et al.* The first proline of PALP motif at the C terminus of presenilins is obligatory for stabilization, complex formation and  $\gamma$ -secretase activities of presenilins. *J. Biol. Chem.* **276**, 33273–33281 (2001).
- Takasugi, N. *et al.* The mechanism of  $\gamma$ -secretase activities through high molecular weight complex formation of presenilins is conserved in *Drosophila melanogaster* and mammals. *J. Biol. Chem.* **277**, 50198–50205 (2002).
- Yu, G. *et al.* Nicastrin modulates presenilin-mediated notch/glp-1 signal transduction and  $\beta$ APP processing. *Nature* **407**, 48–54 (2000).
- Hu, Y., Ye, Y. & Fortini, M. E. Nicastrin is required for  $\gamma$ -secretase cleavage of the *Drosophila* Notch receptor. *Dev. Cell* **2**, 69–78 (2002).
- Edbauer, D., Winkler, E., Haass, C. & Steiner, H. Presenilin and nicastrin regulate each other and determine amyloid  $\beta$ -peptide production via complex formation. *Proc. Natl Acad. Sci. USA* **99**, 8666–8671 (2002).
- Goutte, C., Tsunozaki, M., Hale, V. A. & Priess, J. R. APH-1 is a multipass membrane protein essential for the Notch signaling pathway in *Caenorhabditis elegans* embryos. *Proc. Natl Acad. Sci. USA* **99**, 775–779 (2002).
- Francis, R. *et al.* aph-1 and pen-2 are required for notch pathway signaling,  $\gamma$ -secretase cleavage of  $\beta$ APP, and presenilin protein accumulation. *Dev. Cell* **3**, 85–97 (2002).
- Lee, S. F. *et al.* Mammalian APH-1 interacts with presenilin and nicastrin and is required for intramembrane proteolysis of amyloid- $\beta$  precursor protein and notch. *J. Biol. Chem.* **277**, 45013–45019 (2002).
- Gu, Y. *et al.* APH-1 interacts with mature and immature forms of presenilins and nicastrin and may play a role in maturation of presenilin-nicastrin complexes. *J. Biol. Chem.* **278**, 7374–7380 (2003).
- Steiner, H. *et al.* PEN-2 is an integral component of the  $\gamma$ -secretase complex required for coordinated expression of presenilin and nicastrin. *J. Biol. Chem.* **277**, 39062–39065 (2002).
- Ui, K. *et al.* Newly established cell lines from *Drosophila* larval CNS express neural specific characters. *In Vitro Cell. Dev. Biol. Anim.* **30A**, 209–216 (1994).
- Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
- Fortini, M. E.  $\gamma$ -secretase-mediated proteolysis in cell-surface-receptor signalling. *Nature Rev. Mol. Cell. Biol.* **3**, 673–684 (2002).
- Urban, S. & Freeman, M. Intramembrane proteolysis controls diverse signalling pathways throughout evolution. *Curr. Opin. Genet. Dev.* **12**, 512–518 (2002).
- Fukumoto, H. *et al.* Primary cultures of neuronal and non-neuronal rat brain cells secrete similar proportions of amyloid  $\beta$  peptides ending at A $\beta$ 40 and A $\beta$ 42. *Neuroreport* **10**, 2965–2969 (1999).
- Thinakaran, G. *et al.* Stable association of presenilin derivatives and absence of presenilin interactions with APP. *Neurobiol. Dis.* **4**, 438–453 (1998).
- Morohashi, Y. *et al.* Molecular cloning and characterization of CALP/KChIP4, a novel EF-hand protein interacting with presenilin 2 and voltage-gated potassium channel subunit Kv4. *J. Biol. Chem.* **277**, 14965–14975 (2002).
- Suzuki, N. *et al.* An increased percentage of long amyloid  $\beta$  protein secreted by familial amyloid  $\beta$  protein precursor ( $\beta$ APP717) mutants. *Science* **264**, 1336–1340 (1994).
- Iwatsubo, T. *et al.* Visualization of A $\beta$ 42(43) and A $\beta$ 40 in senile plaques with end-specific A $\beta$  monoclonals: evidence that an initially deposited species is A $\beta$ 42(43). *Neuron* **13**, 45–53 (1994).

**Supplementary Information** accompanies the paper on Nature's website  
(<http://www.nature.com/nature>).

**Acknowledgements** We thank M. Miura, H. Kanuka, T. Igaki, K. Tei, H. Bito, M. Nakafuku and Y. Gotoh for suggestions on S2 and BG2 cells and RNAi assays; G. L. Boulianne, S. Cohen and Y. Ihara for providing Psn cDNA, anti-Smo antibody and anti-C4 antibody, respectively; Takeda Chemical Industries for continuous support to our studies; D.M.A. Mann for comments on the manuscript; and Y. Morohashi, N. Isoo, S. Tanaka and C. Sato for discussions and technical assistance. This work was supported by grants-in-aid from the Ministry of Health and Welfare and the Ministry of Education, Science, Culture and Sports for the 21st Century Center of Excellence programme, Japan. N.T. is a Research Fellow of the Japan Society for the Promotion of Science.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to T.T. (e-mail: taisuke@mol.f.u-tokyo.ac.jp) or T.I. (e-mail: iwatsubo@mol.f.u-tokyo.ac.jp).

# Reactive oxygen species produced by NADPH oxidase regulate plant cell growth

Julia Foreman<sup>\*†</sup>, Vadim Demidchik<sup>†‡</sup>, John H. F. Bothwell<sup>†§</sup>, Panagiotis Mylona<sup>\*</sup>, Henk Miedema<sup>‡</sup>, Miguel Angel Torres<sup>||</sup>, Paul Linstead<sup>\*</sup>, Silvia Costa<sup>\*</sup>, Colin Brownlee<sup>§</sup>, Jonathan D. G. Jones<sup>||</sup>, Julia M. Davies<sup>‡</sup> & Liam Dolan<sup>\*</sup>

<sup>\*</sup> Department of Cell and Developmental Biology, John Innes Centre, Norwich NR4 7UH, UK

<sup>‡</sup> Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK

<sup>§</sup> Marine Biological Association, Citadel Hill, Plymouth PL1 2PB, UK

<sup>||</sup> Sainsbury Laboratory, John Innes Centre, Norwich NR4 7UH, UK

<sup>†</sup> These authors contributed equally to this work

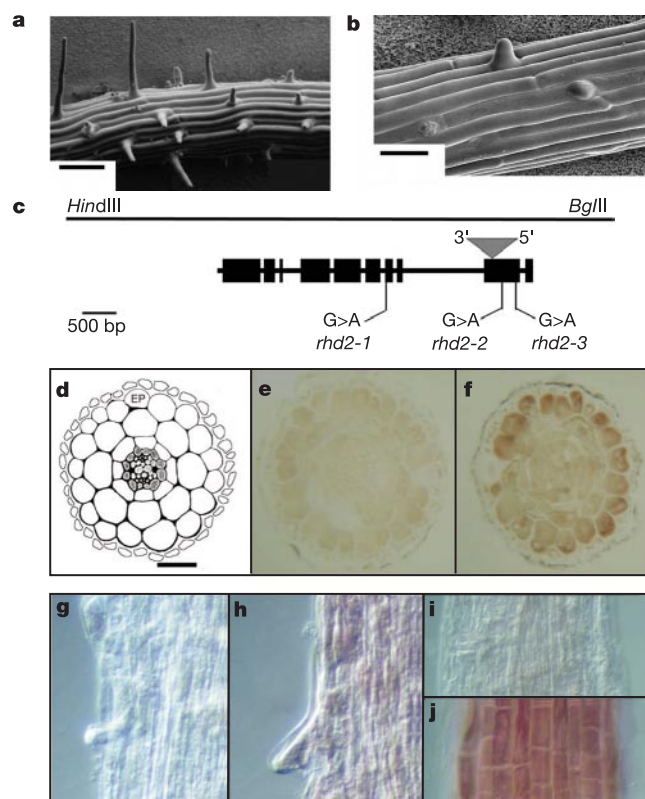
Cell expansion is a central process in plant morphogenesis, and the elongation of roots and root hairs is essential for uptake of minerals and water from the soil.  $\text{Ca}^{2+}$  influx from the extracellular store is required for (and sets the rates of) cell elongation in roots<sup>1</sup>. *Arabidopsis thaliana* *rhd2* mutants are defective in  $\text{Ca}^{2+}$  uptake and consequently cell expansion is compromised—*rhd2* mutants have short root hairs<sup>2,3</sup> and stunted roots. To determine the regulation of  $\text{Ca}^{2+}$  acquisition in growing root cells we show here that RHD2 is an NADPH oxidase, a protein that transfers electrons from NADPH to an electron acceptor leading to the formation of reactive oxygen species (ROS). We show that ROS accumulate in growing wild-type (WT) root hairs but their levels are markedly decreased in *rhd2* mutants. Blocking the activity of the NADPH oxidase with diphenylene iodonium (DPI) inhibits ROS formation and phenocopies *Rhd2*<sup>−</sup>. Treatment of *rhd2* roots with ROS partly suppresses the mutant phenotype and stimulates the activity of plasma membrane hyperpolarization-activated  $\text{Ca}^{2+}$  channels, the predominant root  $\text{Ca}^{2+}$  acquisition system. This indicates that NADPH oxidases control development by making ROS that regulate plant cell expansion through the activation of  $\text{Ca}^{2+}$  channels.

Elevated concentrations of cytoplasmic free calcium ions ( $[\text{Ca}^{2+}]_c$ ) stimulate exocytosis<sup>4</sup> and support elongation of cells in roots<sup>1</sup>.  $\text{Ca}^{2+}$  uptake for root growth is greatest in the elongation zone<sup>5</sup> and is thought to be mediated by plasma membrane hyperpolarization-activated cation channels<sup>6–8</sup>. In roots, once epidermal cells have stopped elongating, tip-growing root hairs elongate from specialized epidermal cells into the surrounding soil. All tip-growing cells (including pollen tubes, fungal hyphae and root hairs) transport  $\text{Ca}^{2+}$  across the plasma membrane into the tip, where the  $[\text{Ca}^{2+}]_c$  is elevated compared with the rest of the cell. In *Arabidopsis* root hairs, the tip-high  $[\text{Ca}^{2+}]_c$  gradient is first observed when slow tip growth initiates at a site on the root-hair bulge<sup>2,9</sup>. Rapidly elongating root hairs maintain the  $[\text{Ca}^{2+}]_c$  gradient until growth ceases when the gradient dissipates<sup>2</sup>. No  $[\text{Ca}^{2+}]_c$  gradient forms in *rhd2* mutants, which develop very short root hairs<sup>2,3</sup> and stunted roots (mean  $\pm$  s.e.m. root length at 8 days: WT,  $25.1 \pm 0.7$  mm; *rhd2*,  $19.9 \pm 0.5$  mm,  $n = 100$ ; significant at the 99.9% level, Student's *t*-test), indicating that RHD2 is required for the establishment of  $\text{Ca}^{2+}$ -mediated cell growth<sup>2</sup> (Fig. 1a and b).

We cloned RHD2 by transposon tagging with the Spm-induced *rhd2-4* allele. DNA flanking the transposon was isolated by inverse polymerase chain reaction (PCR) and sequenced. This showed that the Spm element was inserted into the gene At5g51060 (TIGR;

GenBank accession number AF055355) where it caused a three-base pair duplication of bases 4151–4153 (Fig. 1c). We identified single G to A point mutations in three other mutant alleles, converting tryptophan residues to premature stop codons at positions 597, 737 and 815 in *rhd2-1*, *rhd2-2* and *rhd2-3*, respectively (Fig. 1c). We complemented plants homozygous for the *rhd2-4* mutation with an 8.4-kilobase *Hind*III–*Bgl*II genomic fragment that includes the entire At5g51060 gene, its native promoter and 700 base pairs (bp) of 3' DNA (Fig. 1c). In the T<sub>2</sub> generation, the kanamycin resistance of the T-DNA marker co-segregated with the WT root-hair phenotype. Taken together, these data indicate that the *Rhd2*<sup>−</sup> mutant phenotype is caused by the mutation in the At5g51060 gene. At5g51060 had previously been defined as *Arabidopsis thaliana* respiratory burst oxidase homolog C (*AtrbohC*)<sup>10,11</sup>. This gene hereafter is designated ROOT HAIR DEFECTIVE2/*Arabidopsis thaliana* respiratory burst oxidase homolog C (*RHD2/AtrbohC*). *RHD2/AtrbohC* transcript is present in the epidermis in the proximal regions of the meristem, in the elongation zone, in the differentiation zone and in elongating root hairs (Fig. 1d–j).

*Arabidopsis thaliana* respiratory burst oxidase homologues (*Atrboh*) are homologous to gp91<sup>phox</sup>, the glycosylated transmembrane subunit of the mammalian NADPH oxidase cytochrome that catalyses ROS production<sup>10–12</sup>. There are 10 *Atrboh*s in *Arabidopsis*. The proteins have an approximately 300-residue cytoplasmic amino-terminal region containing two putative EF-hand motifs,



**Figure 1** Characterization of *RHD2/AtrbohC* gene. **a**, WT (scale bar, 100  $\mu$ m). **b**, *rhd2* mutant (scale bar, 50  $\mu$ m). **c**, *RHD2* is located on chromosome 5 on TAC clone K3K7. The intron–exon organization of *RHD2/AtrbohC* and the positions of *rhd2-1*, *rhd2-2*, *rhd2-3* and *rhd2-4* (grey triangle) mutations are shown. **d**, Schematic transverse section, showing the epidermis (EP) (scale bar, 25  $\mu$ m). **e**, *In situ* hybridization of *RHD2/AtrbohC* sense probe (control). **f**, Antisense *RHD2/AtrbohC* probe. **g**, *In situ* hybridization of *RHD2/AtrbohC* sense probe (control). **h**, Antisense *RHD2/AtrbohC* probe in the region where root hairs grow. **i**, *In situ* hybridization of *RHD2/AtrbohC* sense probe (control). **j**, Antisense *RHD2/AtrbohC* probe shows that *RHD2* is expressed in the elongation zone.

<sup>†</sup> Present address: Department of Biology and Curriculum in Genetics and Molecular Biology, CB3280, 108 Coker Hall, University of North Carolina, Chapel Hill, NC 27599-3280, USA.

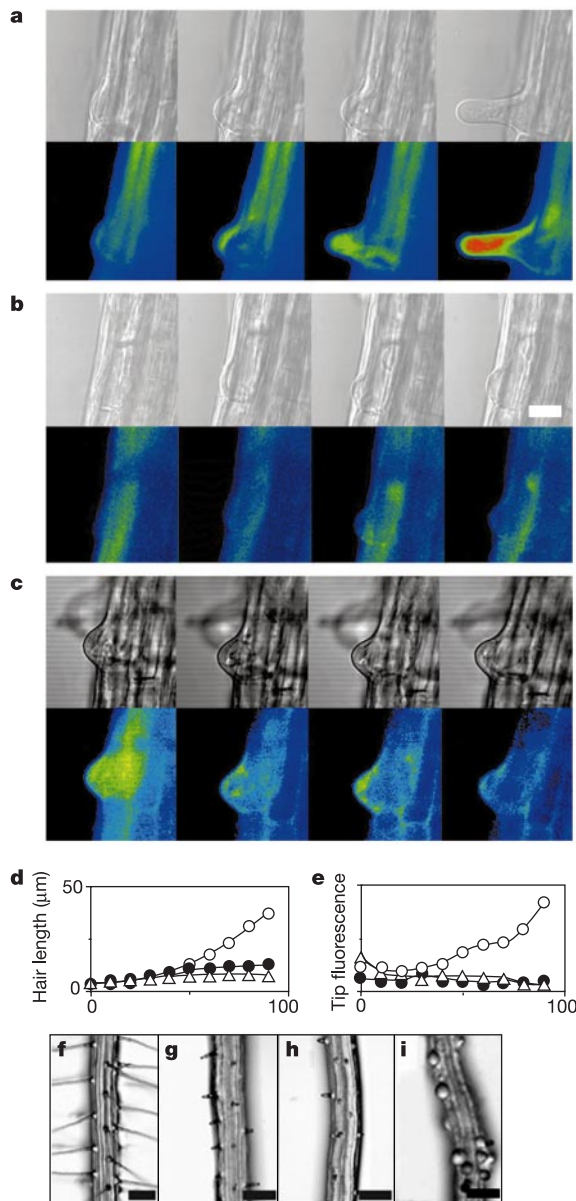


indicating a direct regulatory effect of  $\text{Ca}^{2+}$  ions<sup>10,11</sup>. *AtrbohD* and *AtrbohF* have been shown to produce ROS in response to pathogen infection<sup>13</sup> but so far no developmental roles for *Atrboh*s have been reported.

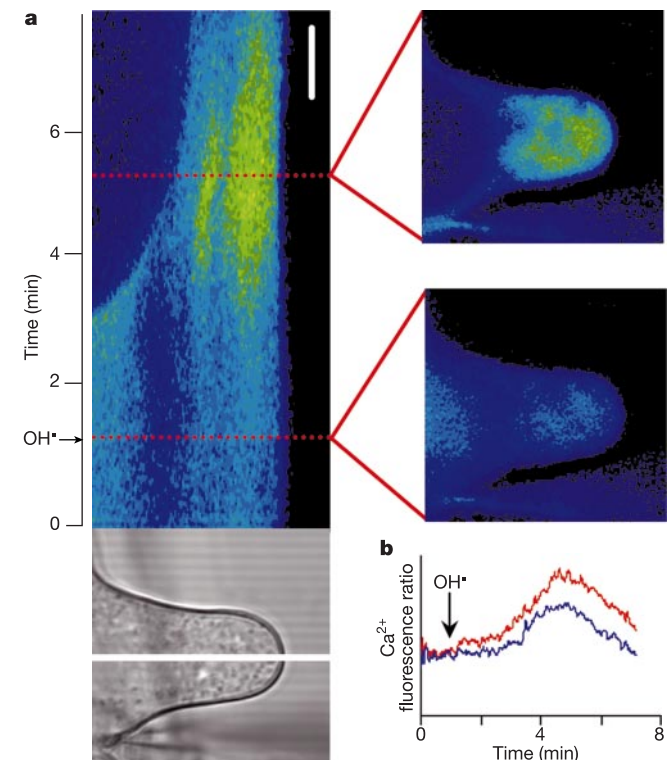
To correlate RHD2/*AtrbohC*-mediated ROS production with growth, a lucigenin chemiluminescence assay of ROS production<sup>14</sup> was used. *rh2* excised root apices (3–4 cm from apex, 10–20 mg fresh weight) produced about half the amount of ROS formed by WT (mean  $\pm$  s.e.m. counts  $\text{s}^{-1} \text{mg}^{-1}$ : WT,  $3.6 \pm 0.6$ ; *rh2*,  $1.54 \pm 0.46$ ;  $n = 4$ ), which is consistent with impaired NADPH oxidase activity in the *rh2* mutant. To image ROS within developing hair cells we ester-loaded 5-(and 6)-chloromethyl-2',7'-

dichlorodihydrofluorescein (CM-H<sub>2</sub>DCF), a ROS-sensitive dye with good intracellular retention into growing roots. Figure 2a shows the increase in ROS within a WT root hair as it emerges initially as a bulge and then elongates by tip growth over the course of an hour. In contrast, there was no ROS accumulation in the *rh2* root-hair bulge over the same period (Fig. 2b; see Supplementary Information for dye-loading control experiments). For WT, both the ROS production at the apex of the root-hair bulge and the elongation rate increased with time ( $n = 8$ ; Fig. 2d and e), whereas *rh2* was impaired in both ROS accumulation and elongation ( $n = 3$ , Fig. 2d and e). Application of 50  $\mu\text{M}$  diphenylene iodonium (DPI), an inhibitor of NADPH oxidase and other flavin-containing enzymes<sup>15</sup>, prevented ROS accumulation in, and elongation of, WT root-hair bulges ( $n = 5$ ; Fig. 2f–h). The fact that treatment of WT with DPI phenocopies *Rhd2*<sup>−</sup> supports the view that an NADPH oxidase is required for root-hair growth. In addition, application of 10  $\mu\text{M}$  DPI to WT inhibited the extension rate of the primary root (measured between 7 and 14 days of growth) by nearly 70% (mean  $\pm$  s.e.m.,  $\text{mm d}^{-1}$ : control,  $3.9 \pm 0.3$ ; DPI-treated root,  $1.2 \pm 0.2$ ;  $n = 10$ ). Taken together, the data indicate that the RHD2/*AtrbohC* protein is required for producing ROS throughout the root and that ROS are required for both root and root-hair growth.

Because *rh2* mutants are defective in setting up the tip-high  $\text{Ca}^{2+}$  gradient in root hairs we propose that ROS produced by the activity of RHD2 are required to stimulate  $\text{Ca}^{2+}$  influx during hair elongation and cell extension<sup>6–8</sup>. It follows that the application of ROS to *rh2* root-hair bulges might elevate  $[\text{Ca}^{2+}]_c$ . During



**Figure 2** ROS accumulation (CM-H<sub>2</sub>DCF imaging) during root-hair elongation: transmission (top) and pseudocolour fluorescent images (bottom) are displayed for WT (**a**), *rh2* (**b**) and WT (**c**) after treatment with DPI. Scale bar in **b**, 20  $\mu\text{m}$  (applies to **a–c**). Red indicates high concentration; blue indicates low concentration. **d**, Increase in root-hair length (mean  $\pm$  s.e.m.); **e**, increase in ROS (arbitrary units) at the hair apex. Open circles, WT, ( $n = 8$ ); filled circles, *rh2* ( $n = 3$ ); open triangles, WT with 50  $\mu\text{M}$  DPI added at zero time ( $n = 5$ ). **f**, WT roots; **g**, *rh2* roots. **h**, WT with 50  $\mu\text{M}$  DPI. **i**, OH<sup>−</sup>-treated *rh2* (Scale bar in **f–i**, 100  $\mu\text{m}$ ).



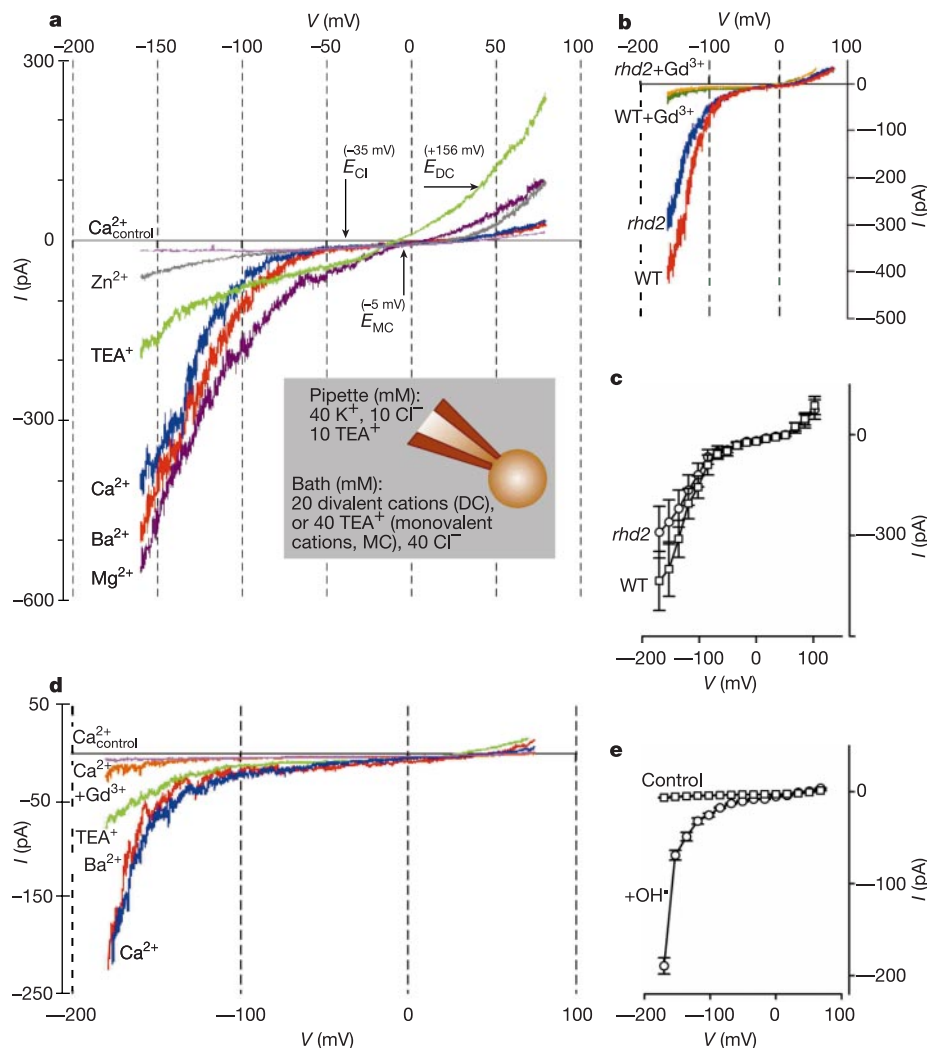
**Figure 3** ROS elevates  $[\text{Ca}^{2+}]_c$  in an *rh2* root-hair bulge. **a**, Left, time course of changes in the calcium green/Texas red pseudocolour  $\text{Ca}^{2+}$  fluorescence ratio for the linescan bar (30  $\mu\text{m}$ ) shown in the transmission image of an *rh2* bulge. Right lower panel, image obtained immediately before the addition of OH<sup>−</sup>; right upper panel, image obtained 3.5 min after addition. Red indicates high concentration; blue indicates low concentration. **b**, Time course of changes  $\text{Ca}^{2+}$  fluorescence ratio in apical (red; 2  $\mu\text{m}$  back from the extreme tip) and sub-apical (blue; 10  $\mu\text{m}$  back from the extreme tip) regions of the bulge. Arrow indicates the time of addition of OH<sup>−</sup>.



endogenous ROS elevation, the superoxide anion produced by a plasma membrane NADPH oxidase can be converted to  $\text{H}_2\text{O}_2$  by superoxide dismutase in the plant apoplast. Hydrogen peroxide (which is known to activate guard-cell  $\text{Ca}^{2+}$  channels<sup>16</sup>) can then be readily converted to hydroxyl radicals ( $\text{OH}^\cdot$ ) in the presence of transition metals (such as  $\text{Cu}^{2+}$  or  $\text{Fe}^{2+}$ )<sup>17,18</sup>.  $\text{OH}^\cdot$  are the most reactive ROS and interact directly with most target biomolecules<sup>18</sup>. Here,  $\text{OH}^\cdot$  were generated by applying a mixture of 2 mM  $\text{H}_2\text{O}_2$ , 0.2 mM  $\text{Cu}^{2+}$  and 0.2 mM ascorbate<sup>18</sup>. As shown in Fig. 3, application of  $\text{OH}^\cdot$  to *rhod2* root-hair bulges caused a rapid elevation in  $[\text{Ca}^{2+}]_c$  ( $n = 4$ ; reported by confocal imaging of calcium green normalized to Texas red). This  $[\text{Ca}^{2+}]_c$  elevation was either delocalized or tip-high (Fig. 3b), which is consistent with  $\text{OH}^\cdot$  treatment's restoring the WT  $[\text{Ca}^{2+}]_c$  gradient. To demonstrate that ROS elevated  $[\text{Ca}^{2+}]_c$  by activating a plasma membrane  $\text{Ca}^{2+}$  influx pathway, spheroplasts were released by laser microsurgery from the apices of WT young root hairs and *rhod2* bulges, then ester-loaded with calcium green-1 acetoxymethyl ester, a non-ratiometric  $\text{Ca}^{2+}$ -reporting fluorophore.  $\text{OH}^\cdot$  again elevated  $[\text{Ca}^{2+}]_c$  but this was prevented by the application of 0.1 mM  $\text{Gd}^{3+}$  (a  $\text{Ca}^{2+}$ -channel

antagonist; data in Supplementary Information). Moreover, increasing ROS levels in *rhod2* roots by treatment with  $\text{OH}^\cdot$  led to depolarized root-hair growth—hairs expanded on all surfaces, leading to the formation of spherical outgrowths (Fig. 2i). The formation of non-polarized hairs upon homogeneous treatment with  $\text{OH}^\cdot$  suggests that the polarized production of ROS is required for the formation of a polarized hair outgrowth.

Because these data suggest that ROS elevate  $[\text{Ca}^{2+}]_c$  by means of  $\text{Ca}^{2+}$  influx we tested for ROS activation of plasma membrane  $\text{Ca}^{2+}$  channels in protoplasts derived from the elongation zone epidermis and in apical spheroplasts isolated from root hairs. With the use of the whole-cell patch clamp configuration, no significant  $\text{Ca}^{2+}$  conductance was evident in WT protoplasts from elongation zone epidermis under control conditions (Fig. 4a). In all protoplasts tested, a  $\text{Ca}^{2+}$ -permeable, inwardly rectifying, hyperpolarization-activated conductance appeared within a few minutes of exposure to the  $\text{OH}^\cdot$ -generating mixture and attained its maximum 20–40 min after exposure ( $n = 4$ ). No activation was achieved by the addition of either 2 mM  $\text{H}_2\text{O}_2$  or 0.2 mM  $\text{Cu}^{2+}$  or 0.2 mM ascorbate individually (data not shown). The  $\text{OH}^\cdot$ -activated conductance was



**Figure 4** Activation of hyperpolarization-activated  $\text{Ca}^{2+}$  channels by ROS. **a**, Whole-cell currents from a WT elongation zone epidermal protoplast. After activation by  $\text{OH}^\cdot$  in 20 mM external  $\text{Ca}^{2+}$ , the protoplast was exposed to different cations in the presence of  $\text{OH}^\cdot$ .  $E$  is the estimated equilibrium potential. **b**,  $\text{OH}^\cdot$ -activated  $\text{Ca}^{2+}$  currents from a WT and an *rhod2* elongation-zone epidermal protoplast. **c**, Mean ( $\pm$ s.e.m.) current-voltage

( $I$ - $V$ ) relationships for elongation-zone  $\text{OH}^\cdot$ -activated  $\text{Ca}^{2+}$  currents;  $n = 4$ . **d**,  $\text{OH}^\cdot$  activation of whole-cell currents from a WT hair apical spheroplast. Protocol and equilibrium potentials as in **a**. **e**, Mean ( $\pm$ s.e.m.)  $I$ - $V$  relationships for WT hair  $\text{OH}^\cdot$ -activated  $\text{Ca}^{2+}$  currents;  $n = 7$ . In **b**, **c** and **e** the bath contained 20 mM  $\text{Ca}^{2+}$ .

readily permeable to divalent cations (Fig. 4a) including  $\text{Ba}^{2+}$ , indicating that this was a  $\text{Ca}^{2+}$  conductance rather than the inwardly rectifying  $\text{K}^{+}$  channel conductance in the root<sup>19</sup>. The  $\text{K}^{+}$  channel blocker tetraethylammonium ( $\text{TEA}^{+}$ ) permeated the  $\text{OH}^{-}$ -activated conductance but  $\text{Zn}^{2+}$  permeation was poor (Fig. 4a), which is consistent with the operation of a  $\text{Ca}^{2+}$  conductance<sup>20</sup>. External  $\text{Gd}^{3+}$  (0.1 mM) blocked the conductance (Fig. 4b). Application of 0.1 and 1 mM  $\text{GdCl}_3$  to WT root apices decreased root elongation rate (4–8-day application and growth interval, mean  $\pm$  s.e.m.,  $\text{mm d}^{-1}$ : control  $10.9 \pm 0.4$ ,  $n = 35$ ; 0.1 mM  $\text{Gd}^{3+}$ ,  $7.4 \pm 0.3$ ,  $n = 52$ ; 1 mM  $\text{Gd}^{3+}$ ,  $4.38 \pm 0.4$ ,  $n = 41$ ), supporting a role for this  $\text{OH}^{-}$ -activated  $\text{Ca}^{2+}$  conductance in growth. Chelation of extracellular  $\text{Ca}^{2+}$  by the apical application of 10 mM BAPTA caused a 10-fold decrease in growth rate (control,  $10.8 \pm 0.3$ ,  $n = 65$ ; BAPTA,  $1.1 \pm 0.1$ ,  $n = 90$ ), confirming  $\text{Ca}^{2+}$  uptake as essential. The  $\text{OH}^{-}$ -induced hyperpolarization-activated  $\text{Ca}^{2+}$  conductance was found in all protoplasts derived from the elongation-zone epidermis of *rhd2* (Fig. 4b and c;  $n = 4$ ). As there was no significant difference in the current amplitudes of the  $\text{OH}^{-}$ -induced conductance between *rhd2* and WT protoplasts (Fig. 4c;  $n = 4$ , recorded in 20 mM  $\text{Ca}^{2+}$ ), the intrinsic sensitivity of the underlying  $\text{Ca}^{2+}$  channel to ROS and the number of channel proteins seem not to be affected by the *rhd2* mutation.

Application of  $\text{OH}^{-}$  (generated in these experiments by 0.5 mM  $\text{Cu}^{2+}$  and 0.5 mM ascorbate<sup>18</sup>) to WT root-hair apical spheroplasts also activated an inwardly rectifying, hyperpolarization-activated conductance in the plasma membrane. This conductance was mediated by  $\text{Ca}^{2+}$ ,  $\text{Ba}^{2+}$  and  $\text{TEA}^{+}$  (Fig. 4d and e;  $n = 7$ ). Activation was apparent within 5 min of the application of  $\text{OH}^{-}$  and attained a maximum within 30 min. No channel activation was observed after the addition of 0.5 mM  $\text{Cu}^{2+}$  alone. The root-hair ROS-activated  $\text{Ca}^{2+}$  conductance was blocked by 0.1 mM  $\text{Gd}^{3+}$  (Fig. 4d), strongly suggesting that this  $\text{Ca}^{2+}$  influx pathway was responsible for the ROS-induced  $\text{Gd}^{3+}$ -sensitive elevation in  $[\text{Ca}^{2+}]_c$  observed in isolated apical spheroplasts (Supplementary Information). Significantly, lanthanides (which block  $\text{Ca}^{2+}$  channels) inhibit root-hair elongation<sup>21</sup>. These data indicate that the defect in ROS production in the *rhd2* mutant compromises the ROS-activated plasma membrane  $\text{Ca}^{2+}$ -channel influx mechanism required for root-hair cell elongation. Our data also indicate that the same mechanism might operate during expansive cell elongation found in other cell types. The involvement of ROS in both types of cell elongation can account for the short root and root-hairless phenotypes characteristic of the *rhd2* mutant.

The demonstration that ROS can activate  $\text{Ca}^{2+}$  channels in the root combined with *rhd2* root-elongation defects indicates that RHD2/*AtrbohC* is part of the mechanism controlling cell expansion. The animal NADPH oxidase gp91<sup>phox</sup> is regulated by Rac proteins, and similar proteins in plants called Rops ('Rho-related GTPases from plants') have been shown to be involved in correct cellular expansion in the root elongation zone, the establishment of the tip-high  $\text{Ca}^{2+}$  gradient and root-hair growth<sup>12,22,23</sup>. We therefore propose that a Rop protein activates RHD2/*AtrbohC* and the resultant ROS (as  $\text{OH}^{-}$ ) then activate the hyperpolarization-activated  $\text{Ca}^{2+}$  channels to facilitate  $\text{Ca}^{2+}$  influx for growth, where it is involved in the modulation of actin dynamics and other processes required for growth. The production of ROS in the growing region of the hair cell indicates that local ROS formation might be part of the mechanism that establishes hair polarity. □

## Methods

### Plants and growth conditions

*Arabidopsis thaliana* (L.) Heyn. lines used in these experiments were derived from the Landsberg *erecta* and Columbia backgrounds. Plants were grown vertically on MS medium (solidified with Phytigel) at 24 °C under continuous illumination. Seedlings were grown to maturity on a 1:1 mix of potting compost (John Innes no. 1) and peat moss.

### Sequencing of mutant alleles

To sequence *rhd2* alleles, the RHD2/*AtrbohC* genomic sequence from WT and mutant plant tissues was amplified by PCR with a mix of *Taq* DNA polymerase (Gibco BRL) and *Pfu* DNA polymerase (Promega). Nucleotide sequences were determined with the ABI PRISM® BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) in conjunction with an Applied Biosystems 3700 DNA sequencer. Sequences were analysed with LASERGENE computer software.

### Inverse PCR and reverse transcriptase-mediated PCR analysis

Genomic DNA was isolated from 6-week-old plants. Inverse PCR was performed as described in ref. 24. *AtrbohC* transcript levels were tested with PCR primers BO1 (5'-GAAGTGTGAGGAGGCTATCACTCTGA-3') and BO2 (5'-GCAAAATCCTTTGAAGTCATTC-3') and were run with the following temperature profile (40 cycles): 2 min at 94 °C, 30 s at 94 °C, 44 s at 55 °C, 1 min at 72 °C. These primers produce a 559-bp fragment when complementary DNA is used as the template and a 640-bp fragment when genomic DNA is used.

### In situ hybridization

A 660-bp fragment of genomic TAC clone K3K7, spanning the first intron and first exon of RHD2/*AtrbohC*, was cloned into pBluescript SK<sup>+</sup>. The antisense probe was transcribed with T3 RNA polymerase; the sense probe was transcribed with T7. *In situ* hybridization of digoxigenin-UTP-labelled RNA probes were performed as described<sup>25</sup>. For *in situ* hybridization of sectioned material, we performed tissue fixation, wax embedding and tissue sectioning as described<sup>25</sup>.

### Constructs and plant transformation

For the complementation of *rhd2* a HindIII–BglII fragment from genomic TAC clone K3K7, containing the full-length RHD2/*AtrbohC* gene, 2,750 bp of its promoter and 700 bp downstream, was cloned into a binary vector. This construct was introduced in *Agrobacterium tumefaciens* strain GV3101 using triparental mating and was transformed into homozygous *rhd2-4* mutant or WT plants by infiltration.

### Sequence analysis

The BLAST search program was used for sequence analysis and comparisons in the Tair, GenBank, EMBL and SwissProt databases<sup>26</sup>. Multiple sequence alignments and a phylogenetic tree based on pairwise dynamic alignment were done with ClustalW (Des Higgins) and TreeViewPPC v. 1.6.2 programs<sup>27,28</sup>. The sequences of *Atrboh*s and gp91<sup>phox</sup> were accessed from GenBank.

### Production of protoplasts and spheroplasts

Root tips were exposed to enzyme solution (1.5% cellulase, 1% cellulysin, 0.1% pectolyase, 0.1% bovine serum albumin, 10 mM KCl, 10 mM  $\text{CaCl}_2$ , 2 mM  $\text{MgCl}_2$ , 2 mM MES, adjusted to pH 5.7 with Tris base, and to 290–300 mosM with sorbitol) for 1–2 h (22 °C) to release elongation-zone epidermal protoplasts (diameter 20  $\mu\text{m}$ ). Apical spheroplasts were released from hairs by laser ablation<sup>7</sup> or by enzymatic digestion of the cell wall (1.5–2 h) of roots from which the tip (containing the elongation zone) had been removed. No release of epidermal protoplasts was observed. Both methods yielded densely cytoplasmic spheroplasts (diameter 12–14  $\mu\text{m}$ ), typical of young root hairs<sup>7</sup>.

### Imaging

Plants 4–5 days old were used, and all fluorescence images were obtained with a Nikon Diaphot 300 inverted microscope coupled to a MRC 1024 scanning-laser confocal system (running Bio-Rad LaserSharp 2.1T). For experiments with CM-H<sub>2</sub>DCF, roots were incubated for 60 min at 4 °C in 5–20  $\mu\text{M}$  CM-H<sub>2</sub>DCF diacetate, acetyl ester<sup>29</sup>, dissolved in DMSO (less than 0.005% final concentration), then washed with 0.1 mM KCl, 0.1 mM  $\text{CaCl}_2$  (pH 6.0) and left for 60 min at 22 °C before experimentation. Dye excitation was at 488 nm; emitted light was detected at 522 nm. Each image consisted of 512 × 512 pixels, representing ten scans of 1 s each, and averaged with Kalman filtering. For cellular measurements of  $[\text{Ca}^{2+}]_c$ , microinjection pipettes were fabricated from 1.5-mm borosilicate pipettes and backfilled with 1 mM calcium green dextran (10 kDa) and 1 mM Texas red dextran (10 kDa). After impalement and recording of a stable membrane potential (measured with a discontinuous current clamp; Axoclamp 2B amplifier from Axon Instruments), pressure was applied (PLI 100 pressure injection system from Medical Systems Corp.) to ~250 kPa. Application of positive potential (+30 to +50 mV) facilitated dye injection by electro-osmotic flow. The impalement pipette was not removed. Cells were bathed in 0.1 mM KCl, 0.1 mM  $\text{CaCl}_2$  (pH 6.0). Calcium green and Texas red fluorescence was monitored at 523 and 606 nm after excitation at 488 and 568 nm, respectively. Image pairs were collected at 1–3-s intervals and analysed with Scion Image software (Scion Corp., Frederick, MD). For cryoscanning electron microscopy, seedlings were placed on moist nitrocellulose paper, mounted on a stub and immersed in liquid nitrogen slush. Roots were transferred to a cold stage. After removal of water by sublimation, roots were sputter coated with gold and observed with a Jeol scanning electron microscope at –147 °C.

### Electrophysiology

The bathing solution comprised (in mM): 20 mM chloride salt of  $\text{Ca}^{2+}$ ,  $\text{Ba}^{2+}$ ,  $\text{Mg}^{2+}$  or  $\text{Zn}^{2+}$ , or 40 mM TEACl with 2 mM MES (adjusted to pH 5.6 with Tris base, and to 290–300 mosM with sorbitol). Pipette solution contained (in mM): 40  $\text{K}^{+}$ , 10  $\text{TEA}^{+}$ , 10  $\text{Cl}^{-}$ , 40 gluconate, 5 BAPTA (free  $\text{Ca}^{2+}$  adjusted to 100 nM with  $\text{Ca}(\text{OH})_2$ ), adjusted to pH 7.2 with 5 mM HEPES/NaOH.  $\text{OH}^{-}$  were generated by the addition of (in mM) either 2  $\text{H}_2\text{O}_2$ , 0.2  $\text{CuCl}_2$ , 0.2 ascorbate or 0.5  $\text{CuCl}_2$ , 0.5 ascorbate<sup>18</sup>. Standard patch clamping procedures were employed<sup>7</sup>. For voltage ramps, the holding voltage was –70 mV; then,

after a 10-s pre-pulse of +80 or +100 mV the voltage was ramped down to -180 mV over 42 s. Data were sampled at 2–5 kHz and low-pass filtered at 1–2 kHz.

Received 18 September 2002; accepted 5 February 2003; doi:10.1038/nature01485.

1. Cramer, G. R. & Jones, R. L. Osmotic stress and abscisic acid reduce cytosolic calcium activities in roots of *Arabidopsis thaliana*. *Plant Cell Environ.* **19**, 1291–1298 (1996).
2. Wymer, C. L., Bibikova, T. N. & Gilroy, S. Cytoplasmic free calcium distributions during the development of root hairs of *Arabidopsis thaliana*. *Plant J.* **12**, 427–439 (1997).
3. Schiefelbein, J. W. & Somerville, C. Genetic control of root hair development in *Arabidopsis thaliana*. *Plant Cell* **2**, 235–243 (1990).
4. Carroll, A. D. *et al.* Ca<sup>2+</sup>, annexins, and GTP modulate exocytosis from maize root cap protoplasts. *Plant Cell* **10**, 1267–1276 (1998).
5. Demidchik, V. *et al.* *Arabidopsis thaliana* root nonselective cation channels mediate calcium uptake and are involved in growth. *Plant J.* **32**, 799–808 (2002).
6. Kiegle, E., Gilliam, M., Haseloff, J. & Tester, M. Hyperpolarisation-activated calcium channels found only in cells from the elongation zone of *Arabidopsis thaliana* roots. *Plant J.* **21**, 225–229 (2000).
7. Véry, A.-A. & Davies, J. M. Hyperpolarization-activated calcium channels at the tip of *Arabidopsis* root hairs. *Proc. Natl Acad. Sci. USA* **97**, 9801–9806 (2000).
8. Miedema, H., Bothwell, J. H. F., Brownlee, C. & Davies, J. M. Calcium uptake by plant cells—channels and pumps acting in concert. *Trends Plant Sci.* **11**, 514–519 (2001).
9. Dolan, L. *et al.* Clonal relationships and cell patterning in the root epidermis of *Arabidopsis*. *Development* **120**, 2465–2474 (1994).
10. Keller, T. *et al.* A plant homolog of the neutrophil NADPH oxidase gp91<sup>phox</sup> subunit gene encodes a plasma membrane protein with Ca<sup>2+</sup> binding motifs. *Plant Cell* **10**, 255–266 (1998).
11. Torres, M. A. *et al.* Six *Arabidopsis thaliana* homologues of the human respiratory burst oxidase (gp91<sup>phox</sup>). *Plant J.* **14**, 365–370 (1998).
12. Segal, A. W. & Abo, A. The biochemical basis of the NADPH oxidase of phagocytes. *Trends Biochem. Sci.* **18**, 43–47 (1993).
13. Torres, M. A., Dangl, J. L. & Jones, J. D. *Arabidopsis* gp91<sup>phox</sup> homologues AtrbohD and AtrbohF are required for accumulation of reactive oxygen intermediates in the plant defence response. *Proc. Natl Acad. Sci. USA* **99**, 517–522 (2002).
14. Williams, A. J. & Cole, P. J. Investigation of alveolar macrophage function using lucigenin-dependent chemiluminescence. *Thorax* **36**, 866–869 (1981).
15. Bolwell, G. P. & Wojtaszek, P. Mechanisms for the generation of reactive oxygen species in plant defence—a broad perspective. *Physiol. Mol. Plant Pathol.* **51**, 347–366 (1997).
16. Pei, Z. M. *et al.* Calcium channels activated by hydrogen peroxide mediate abscisic acid signalling in guard cells. *Nature* **406**, 731–734 (2000).
17. Fry, S. C. Oxidative scission of plant cell wall polysaccharides by ascorbate-induced hydroxyl radicals. *Biochem. J.* **332**, 507–515 (1998).
18. Halliwell, B. & Gutteridge, J. M. C. *Free Radicals in Biology and Medicine* (Oxford Univ. Press, Oxford, 1999).
19. Hirsch, R. E., Lewis, B. D., Spalding, E. P. & Sussman, M. R. A role for the AKT1 potassium channel in plant nutrition. *Science* **280**, 918–921 (1998).
20. Kiss, T. & Osipenko, O. N. Toxic effects of heavy metals on ionic channels. *Pharmacol. Rev.* **46**, 245–267 (1994).
21. Felle, H. & Hepler, P. K. The cytosolic Ca<sup>2+</sup> concentration gradient of *Sinapis alba* root hairs as revealed by Ca<sup>2+</sup>-selective microelectrode tests and fura-dextran ratio imaging. *Plant Physiol.* **114**, 39–45 (1997).
22. Molendijk, A. J. *et al.* *Arabidopsis thaliana* Rop GTPases are localized to tips of root hairs and control polar growth. *EMBO J.* **20**, 2779–2788 (2001).
23. Jones, M. A. *et al.* The *Arabidopsis* Rop2 GTPase is a positive regulator of both root hair initiation and tip growth. *Plant Cell* **14**, 763–776 (2002).
24. Tissier, A. F. *et al.* Multiple independent defective *Suppressor-mutator* transposon insertions in *Arabidopsis*: A tool for functional genomics. *Plant Cell* **11**, 1841–1852 (1999).
25. Coen, E. S. *et al.* *floricaula*: a homeotic gene required for flower development in *Antirrhinum majus*. *Cell* **63**, 1311–1322 (1990).
26. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
27. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
28. Page, R. D. M. TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applic. Biosci.* **12**, 357–358 (1996).
29. Zhang, W. H., Rengel, Z. & Kuo, J. Determination of intracellular Ca<sup>2+</sup> in cells of intact wheat roots: loading of acetoxymethyl ester of Fluo-3 under low temperature. *Plant J.* **15**, 147–151 (1998).

**Supplementary Information** accompanies the paper on Nature's website  
(<http://www.nature.com/nature>).

**Acknowledgements** We thank D. Graham and A. Dark for help with screening and cultivation, respectively; E. Ryan, P. Shaw and K. Roberts for comments on the manuscript; and P. Doerner for support. This work was funded by the BBSRC, the Gatsby Foundation, the Leverhulme Trust and the European Union.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to L.D.  
(e-mail: [liam.dolan@bbsrc.ac.uk](mailto:liam.dolan@bbsrc.ac.uk)).

## Hidden complexity in the mechanical properties of titin

Philip M. Williams\*, Susan B. Fowler†, Robert B. Best†, José Luis Toca-Herrera†‡, Kathryn A. Scott†, Annette Steward† & Jane Clarke†

\* Laboratory of Biophysics and Surface Analysis, School of Pharmaceutical Sciences, University of Nottingham, Nottingham NG7 2RD, UK

† Cambridge University Chemical Laboratory, MRC Centre for Protein Engineering, Lensfield Road, Cambridge CB2 1EW, UK

Individual molecules of the giant protein titin span the A-bands and I-bands that make up striated muscle. The I-band region of titin is responsible for passive elasticity in such muscle<sup>1–4</sup>, and contains tandem arrays of immunoglobulin domains. One such domain (I27) has been investigated extensively, using dynamic force spectroscopy and simulation<sup>5–12</sup>. However, the relevance of these studies to the behaviour of the protein under physiological conditions was not established. Force studies reveal a lengthening of I27 without complete unfolding, forming a stable intermediate that has been suggested to be an important component of titin elasticity<sup>6</sup>. To develop a more complete picture of the forced unfolding pathway, we use mutant titins—certain mutations allow the role of the partly unfolded intermediate to be investigated in more depth. Here we show that, under physiological forces, the partly unfolded intermediate does not contribute to mechanical strength. We also propose a unified forced unfolding model of all I27 analogues studied, and conclude that I27 can withstand higher forces in muscle than was predicted previously.

Titin's I-band region is largely composed of tandem arrays of immunoglobulin domains along with the PEVK region, believed to be largely unstructured<sup>13–16</sup>. Most studies suggest that under physiological forces the increase in the length of titin derives from straightening of the immunoglobulin regions and unfolding of the PEVK domains<sup>3,17</sup>, although it has also been suggested that, at extremes of force, a small number of the immunoglobulin domains could unfold<sup>18,19</sup>. The resistance of these immunoglobulin domains to force has been studied extensively by single-molecule studies<sup>5,10,20–22</sup> that mostly used atomic force microscopy (AFM) techniques. They have been shown to be the most mechanically stable proteins studied to date, at the high loading rates of the experimental regime. However, extrapolation of these data to low loading rates that are likely to be physiologically relevant suggests that there will be considerable unfolding of the immunoglobulin domains. This is hard to reconcile with results that suggest that these domains are very stable in the intact sarcomere, and that they are likely to unfold very rarely *in vivo*<sup>3,17,23</sup>. To understand fully how the immunoglobulin domains resist force, it is important to investigate in detail how the unfolding energy landscape responds to the application of force—in particular, to ask how it will respond to the low forces relevant to physiological conditions.

I27 (Fig. 1) unfolds through an intermediate (I) under force in which the A-strand is detached. This intermediate was first observed in modelling studies<sup>11</sup>, confirmed by careful analysis of AFM traces<sup>6</sup>, and has been characterized by biophysical and structural techniques<sup>9</sup>. However, it is not observed in chemical denaturation studies<sup>24</sup>. Force versus extension curves of wild-type (WT) I27 indicate that molecular rearrangement to I occurs around 100 pN (ref. 6), and as at AFM loading rates the WT unfolds above this force, the dynamic force spectrum for WT I27 represents the I to U transition (Fig. 2a): that is, this intermediate is the dominant species

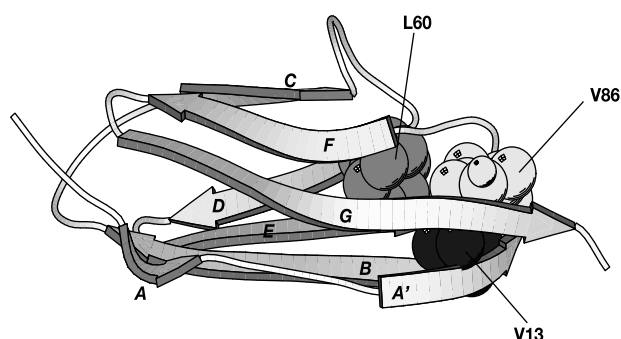
‡ Present address: Center for Ultrastructure Research, Universität für Bodenkultur Wien, Gregor Mendel Strasse 33, A-1180 Vienna.



in the AFM experiments. The AFM data indicate that the barrier to unfolding, at the speeds used in the AFM, is situated  $\sim 3 \text{ \AA}$  along the force trajectory. This experimentally determined unfolding distance ( $x_u$ ) is that between I and the main transition state, and the unfolding rate extrapolated to zero force ( $k_u^0$ ) determined from the AFM data is the rate of unfolding of I (ref. 9).

To study the molecular origins of mechanical stability, single conservative amino-acid mutations have been used to probe the structure of the force transition state (Table 1)<sup>25</sup>. Force behaviour unaffected on mutation indicates that the mutation destabilizes the transition state to the same extent as the 'ground' state (the activation energy to unfold and the distance to the transition state are unchanged). For example, mutation of the leucine at position 60 to alanine (L60A), which destabilizes the native state by  $>5 \text{ kcal mol}^{-1}$ , does not alter the dynamic force spectrum (Fig. 3a, triangles). Hence this deeply buried residue in the E-strand of the protein remains completely structured in the transition state under force. The mutation V13A (change in free energy of unfolding on mutation,  $\Delta\Delta G \approx 2.4 \text{ kcal mol}^{-1}$ ), however, lowers the unfolding forces, implying that the mutation destabilizes the intermediate more than the transition state, meaning that the A'-strand is partially disrupted in the transition state<sup>25</sup>. In both L60A and V13A, the position of the transition state,  $x_u$  (reflected in the pulling speed dependence of the unfolding force), remains the same as WT.

The mutation V86A, however, markedly alters the dynamic force spectrum, with both a reduction in force and an increase in  $x_u$  (Table 1). The change in  $x_u$  is confirmed by the associated narrowing of the force histogram (Fig. 3b inset). Remarkably, there is an increase in the apparent kinetic stability on mutation as the extrapolated unfolding rate reduces 10-fold. Similar effects have been seen with proline mutants V11P, V13P and V15P (ref. 7). It was suggested that the change in  $k_u$  in these proline mutants occurred because the proline mutants stabilized I27 (ref. 7). This would be a surprising result, and indeed we show here that these mutants are, in fact, significantly destabilized (Table 1). How could such highly destabilizing mutations result in a large decrease in the rate of unfolding? Comparison with the L60A and V13A mutants show that this anomalous behaviour is not a simple effect of the large loss of stability upon mutation (L60A is the most destabilized mutant) or of location (V13 is in the same region as the V11, V15 and V86). It has previously been suggested that an observed increase in  $x_u$  can be ascribed to movement of the transition state, away from the native state<sup>7</sup>. However, this simple explanation does not explain the paradoxical decrease in  $k_u^0$ . Here, we suggest that these mutants unfold before force stabilizes the intermediate, and the dynamic force spectrum reflects unfolding from the N to the major transition



**Figure 1** Ribbon diagram of I27, showing the  $\beta$ -strands labelled A–G and the location of the mutations described. In the force experiments a molecule containing eight tandem repeats of this domain is pulled<sup>26</sup>. This figure was created from Protein Data Bank (http://www.rcsb.org) structure entry 1tit; ref. 30.

**Table 1** Thermodynamic and kinetic parameters for wild-type and mutant TI I27

| Mutant    | $\Delta\Delta G_{D-N}$<br>(kcal mol <sup>-1</sup> ) | $k_u^0$<br>(s <sup>-1</sup> )   | $x_u$<br>( $\text{\AA}$ ) |
|-----------|-----------------------------------------------------|---------------------------------|---------------------------|
| Wild type |                                                     | $(2.9 \pm 1.1) \times 10^{-4}$  | $3.2 \pm 0.1$             |
| V13A      | $2.4 \pm 0.1^*$                                     | $(8.4 \pm 1.6) \times 10^{-4}$  | $3.4 \pm 0.2$             |
| L60A      | $5.3 \pm 0.1^*$                                     | $(9.6 \pm 19.1) \times 10^{-4}$ | $3.0 \pm 0.7$             |
| V86A      | $4.9 \pm 0.1^*$                                     | $(3.8 \pm 3.2) \times 10^{-5}$  | $5.4 \pm 0.1$             |
| V11P      | $4.1 \pm 0.1$                                       | $7 \times 10^{-6}\dagger$       | $5.0\dagger$              |
| V13P      | $4.3 \pm 0.1$                                       | $8 \times 10^{-7}\dagger$       | $6.0\dagger$              |
| V15P      | $2.2 \pm 0.1$                                       | $7 \times 10^{-6}\dagger$       | $4.5\dagger$              |

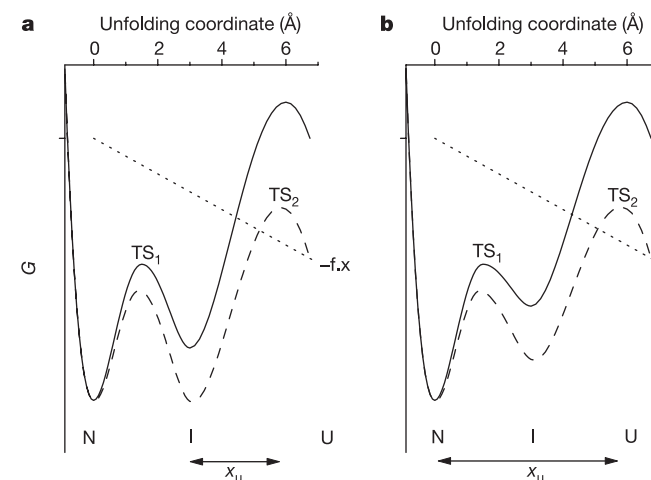
$\Delta\Delta G_{D-N}$ , change in free energy upon mutation determined using equilibrium denaturation ( $\pm$  standard error).  $k_u^0$  and  $x_u$  are determined from two-state analysis as described in Methods ( $\pm$  standard deviation).

\*Data taken from ref. 9. Value is the  $\Delta\Delta G$  in a model unfolding intermediate.

†Data taken from ref. 7. No estimation of errors was given in ref. 7, but we assume they were of the same order as those described above.

state TS<sub>2</sub> (Fig. 2b). The  $x_u$  for these mutants represents the distance between N and TS<sub>2</sub>. As now the extrapolated unfolding rate is the rate of the unfolding of N, it is not comparable to the rate of unfolding of I determined for WT. All AFM measurements of WT unfolding have thus measured the relative height and position of the barrier to unfolding not from the native state but from the intermediate. Modelling the unfolding kinetics over multiple transition states can test this scheme.

A three-state model was established (Fig. 2, Methods and Supplementary Information) where the initial distance from I to TS<sub>2</sub> was set to match the WT data of  $3.2 \text{ \AA}$ , and the distance from N to TS<sub>2</sub> was set to  $5.4 \text{ \AA}$  from the V86A data. The transition from N to I, which corresponds to an initial estimated distance of  $2.2 \text{ \AA}$ , was set to occur at  $100 \text{ pN}$ . This means that the energy difference between N and I is  $3.2 \text{ kcal mol}^{-1}$  (because at equilibrium the energy put into the system is  $2.2 \times 100 \text{ pN \AA}$ , and  $1 \text{ kcal mol}^{-1} = 69 \text{ pN \AA}$ ). The initial unfolding rate from I over TS<sub>2</sub> was set at  $3 \times 10^{-4} \text{ s}^{-1}$  from the WT data, and the ratio of the unfolding and refolding rate constants between N and I set from the  $\Delta\Delta G$  of  $3.2 \text{ kcal mol}^{-1}$ . The model was iterated until all the experimental data (WT, V13A and V86A) could be reproduced by changes in rate constant alone



**Figure 2** The dynamical transition N to I in the unfolding free-energy diagram of I27. The dotted lines ( $-f \cdot x$ ) indicate the change in energy owing to applied force, transforming the original landscape (solid line) to that under force (dashed line). **a**, WT I27 unfolds via an intermediate state (I) defined by two transition states, TS<sub>1</sub> and TS<sub>2</sub>. The closeness of I in energy to the native state (N), and the large transition state TS<sub>2</sub> compared to TS<sub>1</sub> (exaggerated for illustration), creates the conditions where force can deform the landscape (dashed) so that I is populated before unfolding occurs over TS<sub>2</sub>. Dynamic force spectroscopy of WT reveals the displacement of TS<sub>2</sub> from I ( $x_u$ ), and the unfolding kinetics of I to U. **b**, Under the experimental range of loading rates, V86A unfolds before the dynamical transition occurs. Dynamic force spectroscopy of V86A reveals the displacement of TS<sub>2</sub> from N, and the unfolding kinetics of N to U.

(Fig. 3b) and validated by the agreement between predicted and measured force distributions (Fig. 3b inset).

Our model predicts the distance between I and TS<sub>2</sub> to be 3.0 Å. This is slightly shorter than that predicted from a two-state fit, as the three-state model allows some unfolding from N (Supplementary Information). The N to TS<sub>2</sub> distance is 5.9 Å and, for the same reason, slightly longer than the two-state prediction for V86A. Our model predicts that above 100 pN, the WT unfolds from the intermediate structure, which is 3.8 kcal mol<sup>-1</sup> less stable than the native structure. Below these forces, therefore, WT will unfold from the native state, and the dynamic force spectrum for unfolding will have the force scale of V86A representing the 5.9 Å barrier displacement. It is, therefore, possible to predict the force spectrum for the WT below the 100 pN dynamical transition force (Fig. 3b) which leads to an extrapolated unfolding rate at zero force of 10<sup>-6</sup> s<sup>-1</sup>; some 300 times slower than estimated previously. Importantly, therefore, I27 is able to withstand forces below 100 pN considerably longer than was thought (see hatched region in Fig. 3b).

Under the forces tested, V86A does not undergo the dynamical transition from N to I, and the I state under force remains above the N primary minimum (Fig. 2b). The highest force of 142 pN indicates that the I state, at 2.9 Å, is at least 6 kcal mol<sup>-1</sup> above N. The unfolding rate of V86A (from N) is determined to be 3 × 10<sup>-5</sup> s<sup>-1</sup>, which is significantly faster than the rate of unfolding of WT from N. These rates correspond to ΔΔG<sub>TS-N</sub> of 2 kcal mol<sup>-1</sup>

and give a Φ-value of ~0.4, similar to that of V13A that is in the same region of the protein (Φ = 1 - (ΔΔG<sub>TS-N</sub>/ΔΔG<sub>I-N</sub>))<sup>25</sup>. The data also indicate that the V13A mutation destabilizes N and I equally. For a mutant to unfold from N, such as V86A, and not from I, such as V13A, the mutation must destabilize I more than N and/or lower the transition state barrier sufficient so that unfolding occurs before the intermediate can form. Thus very destabilizing mutants in regions that remain structured in the intermediate and transition state, such as L60A, do not affect the unfolding behaviour. V13P is more destabilizing than V13A, and is likely to be structurally disruptive.

It has been recently suggested that mechanical properties derived from AFM experiments can be used to predict the behaviour of titin molecules at physiological forces, estimated variously to be around 10 pN (ref. 10). From our model, we are now in a position to describe the mechanical properties of this domain under a range of forces. Above 100 pN, the unfolding behaviour is that which has been probed by the force experiments, and this is dictated by the stability of the unfolding intermediate. At forces below 100 pN, which have not been investigated by AFM, the strength of the domain is greater than that predicted by direct extrapolation, and under these physiologically more relevant conditions the intermediate, which is not populated, plays no role in determining the mechanical strength. The rate constant for unfolding in the absence of force of 5 × 10<sup>-4</sup> s<sup>-1</sup> (compared to the forced unfolding rate shown here to be 10<sup>-6</sup> s<sup>-1</sup>) confirms that a switch in unfolding trajectory must occur at some level of force<sup>8,9</sup>. This difference in rate constants corresponds to a difference in transition state energy of ~4 kcal mol<sup>-1</sup>; so, making the reasonable assumption that the zero-force transition state is unaffected by force, the solution pathway will dominate at all forces below 43 pN.

AFM is limited in its application to the study of unfolding processes owing to the limited range of loading rates available. Through the analysis of the behaviour of conservative mutations that shift the unfolding rate it has been possible to extend the dynamic range of AFM experimentation. Our studies reveal that I27 has three different responses to applied force. At zero and low loading rates, the principal barrier to unfolding is that characterized by the denaturant studies, associated with a  $k_u^0$  of 5 × 10<sup>-4</sup> s<sup>-1</sup>. This transition state is not probed by force experimentation. As the loading increases, a switch in the unfolding pathway occurs across a second barrier, described here, with a  $k_u^0$  of 10<sup>-6</sup> s<sup>-1</sup>. As the loading increases further, the protein populates an intermediate, which increases the resistance of the protein to further increases in force, but still unfolds over this second barrier. This complex energy landscape emphasizes the dangers inherent in extrapolation of mechanical behaviour from experimental to physiological conditions. The landscape remains to be fully characterized by exploiting both protein engineering and alternative biophysical methods. □

## Methods

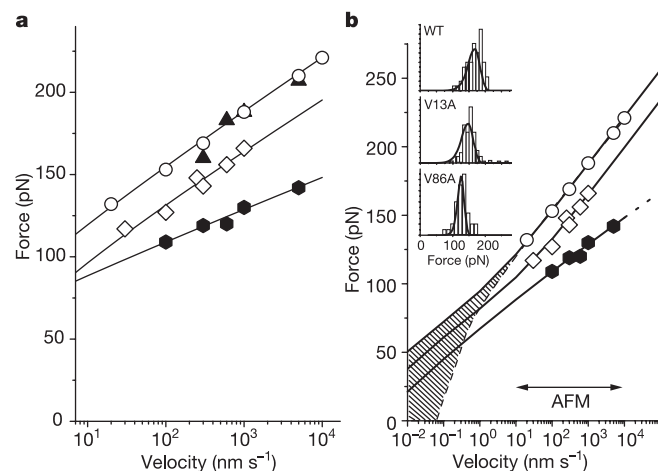
### Experiments

All experiments were performed on protein constructs with eight tandem repeats of TI I27, either wild-type or mutant. All mutant proteins were fully folded and stable under the experimental conditions, as judged by a number of biophysical techniques including fluorimetry and NMR spectroscopy (Data not shown)<sup>24</sup>. The proteins were constructed and purified as described<sup>26</sup>. The AFM experiments were performed at pH 7.4 at ambient temperature, using an Asylum Research molecular force probe (Santa Barbara), as described previously<sup>27</sup>. The stability of the mutant protein (single modules) was determined by equilibrium denaturation as described<sup>24</sup>; we have shown that I27 has the same stability in the multimeric construct<sup>27</sup>.

### Data analysis and modelling

All rupture events except the first peak in each test were used to form the mean unfolding force for each retract speed. Initial estimates of  $k_u^0$  and  $x_u$  (and those in Table 1) were obtained by fitting an analytical approximation of the unfolding of the two-state model (Supplementary Information)<sup>28</sup>. Errors were estimated using the bootstrap method<sup>29</sup>.

The unfolding forces of the three-state system were modelled by solving the rate equations of the time evolution of the three states under a nonlinear ramp of force using numerical computation (Supplementary Information)<sup>28</sup>. (Note that this simplest three-



**Figure 3** Unfolding kinetics of WT I27 and mutants. **a**, A plot of the mean unfolding force versus pulling speed for WT (circles) and mutants L60A (triangles), V13A (diamonds) and V86A (hexagons). V86A exhibits a significant difference of slope (solid lines are fits of a two-state analytical approximation as described in Methods. Data for WT, L60A and V13A are taken from ref. 25). **b**, Modelling of the three-state system with the dynamical transition occurring at 100 pN reveals the true pulling speed dependence of WT I27 and mutants. V86A unfolds over the same transition state as WT and the other mutants presented, but from the native and not the intermediate state. The AFM is limited in the range of retract velocities that may be used, and the dynamic force spectrum for V86A cannot be predicted at high rates of loading (dotted continuation). There is a region in the force spectrum (shaded) where WT I27 can withstand forces for considerably longer than was previously thought (the dashed line is the predicted dynamic force spectrum based on the original two-state unfolding process<sup>26</sup>). Inset, the force histograms for WT, V13A and V86A recorded at 300 nm s<sup>-1</sup>, and those predicted by the model at this speed. Note that the model would also predict a narrowing of the distribution for V86A, which should be reflected in a lower standard deviation. However, the measured distributions will always be widened through noise (estimated to be ~20 pN). The variance of the distribution for each unfolding event will be equal to the variance in force (4.11/ $x_u$ )<sup>2</sup> plus the variance in noise (20)<sup>2</sup>, predicting that WT should have a standard deviation of around 24 pN and V86A around 21 pN. The experimental values are WT 23 pN and V86A 21 pN at 300 nm s<sup>-1</sup>. This is as predicted, but the results are the same, within error, as they are dominated by noise.

state model assumes that the relatively conservative V86A mutation does not cause an abrupt switch of unfolding pathway to an alternative transition state. The wealth of information from protein folding experiments using mutants would suggest that this alternative is almost without precedent, but it cannot be entirely ruled out.)

Numerical methods are used in preference to the more commonly employed Monte Carlo method, as the full and continuous distributions of unfolding forces are revealed. The results from these two methods are shown to be identical (Supplementary Information).

Received 10 December 2002; accepted 20 February 2003; doi:10.1038/nature01517.

1. Furst, D. O., Osborn, M., Nave, R. & Weber, K. The organization of titin filaments in the half-sarcomere revealed by monoclonal-antibodies in immunoelectron microscopy—A map of 10 nonrepetitive epitopes starting at the Z-line extends close to the M-line. *J. Cell Biol.* **106**, 1563–1572 (1988).
2. Labeit, S., Gautel, M., Lakey, A. & Trinick, J. Towards a molecular understanding of titin. *EMBO J.* **11**, 1711–1716 (1992).
3. Linke, W. A. *et al.* Towards a molecular understanding of the elasticity of titin. *J. Mol. Biol.* **261**, 62–71 (1996).
4. Linke, W. A., Stockmeier, M. R., Ivemeyer, M., Hosser, H. & Mundel, P. Characterizing titin's I-band Ig domain region as an entropic spring. *J. Cell Sci.* **111**, 1567–1574 (1998).
5. Carrion-Vazquez, M. *et al.* Mechanical and chemical unfolding of a single protein: A comparison. *Proc. Natl Acad. Sci. USA* **96**, 3694–3699 (1999).
6. Marszalek, P. E. *et al.* Mechanical unfolding intermediates in titin modules. *Nature* **402**, 100–103 (1999).
7. Li, H. B., Carrion-Vazquez, M., Oberhauser, A. F., Marszalek, P. E. & Fernandez, J. M. Point mutations alter the mechanical stability of immunoglobulin modules. *Nature Struct. Biol.* **7**, 1117–1120 (2000).
8. Brockwell, D. J. *et al.* The effect of core destabilization on the mechanical resistance of I27. *Biophys. J.* **83**, 458–472 (2002).
9. Fowler, S. B. *et al.* Mechanical unfolding of a titin Ig domain: Structure of unfolding intermediate revealed by combining AFM, molecular dynamics simulations, NMR and protein engineering. *J. Mol. Biol.* **322**, 841–849 (2002).
10. Li, H. B. *et al.* Reverse engineering of the giant muscle protein titin. *Nature* **418**, 998–1002 (2002).
11. Lu, H., Israelowitz, B., Krammer, A., Vogel, V. & Schulten, K. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.* **75**, 662–671 (1998).
12. Lu, H. & Schulten, K. The key event in force-induced unfolding of titin's immunoglobulin domains. *Biophys. J.* **79**, 51–65 (2000).
13. Labeit, S. & Kolmerer, B. Titins—Giant proteins in charge of muscle ultrastructure and elasticity. *Science* **270**, 293–296 (1995).
14. Trinick, J. Cytoskeleton—Titin as a scaffold and spring. *Curr. Biol.* **6**, 258–260 (1996).
15. Tskhovrebova, L. & Trinick, J. Direct visualization of extensibility in isolated titin molecules. *J. Mol. Biol.* **265**, 100–106 (1997).
16. Ma, K., Kan, L. S. & Wang, K. Polyproline II helix is a key structural motif of the elastic PEVK segment of titin. *Biochemistry* **40**, 3427–3438 (2001).
17. Linke, W. A. & Granzier, H. A spring tale: New facts on titin elasticity. *Biophys. J.* **75**, 2613–2614 (1998).
18. Minajeva, A., Kulke, M., Fernandez, J. M. & Linke, W. A. Unfolding of titin domains explains the viscoelastic behavior of skeletal myofibrils. *Biophys. J.* **80**, 1442–1451 (2001).
19. Tskhovrebova, L. & Trinick, J. Flexibility and extensibility in the titin molecule: Analysis of electron microscope data. *J. Mol. Biol.* **310**, 755–771 (2001).
20. Tskhovrebova, L., Trinick, J., Sleep, J. A. & Simmons, R. M. Elasticity and unfolding of single molecules of the giant muscle protein titin. *Nature* **387**, 308–312 (1997).
21. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M. & Gaub, H. E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109–1112 (1997).
22. Kellermayer, M. S. Z., Smith, S. B., Granzier, H. L. & Bustamante, C. Folding-unfolding transitions in single titin molecules characterized with laser tweezers. *Science* **276**, 1112–1116 (1997).
23. Gautel, M., Lehtonen, E. & Pietruschka, F. Assembly of the cardiac I-band region of titin/connectin: Expression of the cardiac-specific regions and their structural relation to the elastic segments. *J. Muscle Res. Cell Motil.* **17**, 449–461 (1996).
24. Fowler, S. B. & Clarke, J. Mapping the folding pathway of an immunoglobulin domain: Structural detail from phi value analysis and movement of the transition state. *Structure* **9**, 355–366 (2001).
25. Best, R. B., Fowler, S. B., Toca-Herrera, J. L. & Clarke, J. A simple method for probing the mechanical unfolding pathway of proteins in detail. *Proc. Natl Acad. Sci. USA* **99**, 12143–12148 (2002).
26. Steward, A., Toca-Herrera, J. L. & Clarke, J. Versatile cloning system for construction of multimeric proteins for use in atomic force microscopy. *Protein Sci.* **11**, 2179–2183 (2002).
27. Best, R. B., Li, B., Steward, A., Daggett, V. & Clarke, J. Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation. *Biophys. J.* **81**, 2344–2356 (2001).
28. Evans, E. Probing the relation between force—lifetime—and chemistry in single molecular bonds. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 105–128 (2001).
29. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge Univ. Press, Cambridge, UK, 1992).
30. Improt, S., Politou, A. S. & Pastore, A. Immunoglobulin-like modules from titin I-band: Extensible components of muscle elasticity. *Structure* **4**, 323–337 (1996).

**Supplementary Information** accompanies the paper on Nature's website  
(<http://www.nature.com/nature>).

**Acknowledgements** P.M.W. is an EPSRC Advanced Research Fellow, and J.C. is a Wellcome Trust Senior Research Fellow. Funding was provided by the EPSRC, the MRC and the Wellcome Trust. S.B.F. was supported by the EPSRC and Newnham College, and R.B.B. by the Cambridge Commonwealth Trust.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to J.C.  
(e-mail: [jc162@cus.cam.ac.uk](mailto:jc162@cus.cam.ac.uk)).



# Bench marks

Everyday equipment for routine analysis.



## Chemically inert bottle caps

**Omnifit**

[www.omnifit.com](http://www.omnifit.com)

*If the cap fits*

These chemically inert PTFE bottle caps can be used with aggressive clean-in-place solutions, as well as for connecting tubing to inert fluidic systems, solvent reservoirs or waste bottles. Available in standard GL45 and GL32 thread sizes, with other sizes available on request, the bottle caps come in three basic designs, depending on requirements. A 'dual-connection' design is ideal for applications requiring a variety of tubing types and/or sizes to be connected to a single cap, while the 'no-twist' design has a central independently rotating PTFE block into which tubing lines are connected. The third option, a 'low profile' design ideal for limited space, is supplied fitted with 1-metre lengths of PTFE tubing already connected to the cap.

## Microlute

**Porvair Sciences**

[www.porvair-sciences.com](http://www.porvair-sciences.com)

*Manifold benefits*

Porvair's new acrylic manifold for the Microlute microplate system allows total automation of high-throughput solid phase extraction or protein precipitation experiments. Compatible with most commercial robotic control systems, the manifold permits robotic manipulators to rapidly and simply access and lift just the upper part of the Microlute. Each manifold comes complete with an adapter to fit standard 1 ml plates, waste collection reservoir tray and a square well collection plate.

## Genevac Model DD-4X

**Genevac**

[www.genevac.com](http://www.genevac.com)

*General-purpose centrifugal solvent evaporation*

Improved speed of evaporation, shorter condenser defrost time, more powerful software (including remote control functions) and the

flexibility to accommodate a wide range of sample formats are key benefits claimed for the new Model DD-4X. A redesigned centrifuge chamber, together with powerful IR lamps, dramatically increases the speed of evaporation. To improve high-throughput capability the waste solvent condenser chamber volume on the DD-4X has been raised and the condenser power has been increased by 50%. Interchangeable high-performance, solid aluminium rotors offer high-capacity handling of test tubes and vials. Alternatively, a swing rotor will accommodate tubes, vials, microtitre plates and Fast-Stack multi-deck microplate holders.

## Pipetboy Comfort

**Integra Biosciences Group**

[www.integra-biosciences.com](http://www.integra-biosciences.com)

*Lightweight ergonomic pipette*

Weighing only 190 grams, the new Pipetboy Comfort is designed for comfortable single-handed pipetting in the 1–100 millilitre volume range. The Pipetboy Comfort has in-built features to prevent accidental misuse. Incorporating a sterile PTFE-membrane filter the

Pipetboy Comfort prevents contamination from overfilling as the penetration of liquid is prevented. In addition the Pipetboy comfort will not squirt out under pressure even if a filled pipette is unintentionally aspirated. For sensitive samples all parts of the Pipetboy Comfort that come into contact with liquid can be autoclaved clean at 121 °C for 30 minutes. It also comes with a combined all-in-one stand and charging station that can be wall-mounted or can sit on the laboratory bench.

*These notes are compiled in the Nature office from information provided by the manufacturers.*

## Contacts

**Publisher:** Ben Crowe

**Editor:** Paul Smaglik

**Marketing Manager:** David Bowen

### European Head Office, London

The Macmillan Building  
4 Crinan Street  
London N1 9XW, UK  
Tel +44 (0) 20 7843 4961  
Fax +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**Senior European Sales Manager:**  
Nevin Bayoumi (4978)

### UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

### Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

**Scandinavia:** Silje Opstrup (4994)

### France/ Belgium:

Amelie Pequignot (4974)

**Production Manager:** Billie Franklin

To send materials use London  
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

### International

**Advertising Coordinator:**

Hind Berrada (4935)

**Naturejobs web development:**

Tom Hancock

**Naturejobs online production:**

Ben Lund

### European Satellite Office

**Germany/ Austria/ Switzerland:**

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/2

Fax + 49 89 54 90 57 20

e-mail: [p.phelan@nature.com](mailto:p.phelan@nature.com)

[o.wulffen@nature.com](mailto:o.wulffen@nature.com)

### US Head Office, New York

345 Park Avenue South,  
10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**US Sales Manager:** Peter Bless

**US Advertising Coordinator:**

Linda Adam

### Japan Head Office, Tokyo

MG Ichigaya Building (5F),  
19-1 Harakakamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

**Asia-Pacific Sales Director:**

Hideki Watanabe

e-mail: [h.watanabe@nature.com](mailto:h.watanabe@nature.com)

# naturejobs

## You can't go home again

**C**ontrary to the title of Thomas Wolfe's novel, you can go home. It is just particularly difficult if you're a scientist, and even more so if you're from a country such as Italy or Spain where local patronage plays a big part in career advancement. But several scientists I visited in Italy this month have managed to leave and return — despite Italy's turbulent science-funding history.

Because they realize they're the exception, they all urge young Italian scientists to go away first. "I'm trying to pass an internal rule," says Pietro Calissano, of the CNR's Institute of Neurobiology and Molecular Medicine in Rome, whose CV covers the United States, Britain and Israel. "As soon as you get your PhD in this institute you must go abroad."

But getting back can be tricky. Sometimes pure luck steps in. Brunella Franco left in 1994 to do a postdoc at Baylor University in the United States under Andrea Ballabio, with whom she had studied in Naples. Ballabio went on to become the scientific director of the Telethon Institute of Genetics and Medicine in Milan. When it moved to Naples in 2000, both Neapolitans found themselves back home.

Lino Polito, director of the Institute of Genetics and Biophysics in Naples, took no chances. In 1971, he left Italy to work with Ed Southern at the University of Edinburgh, UK — but with a guarantee of a position in Italy when the job ended.

The current funding climate in many European countries, including Italy, is not encouraging young scientists to stay put. And schemes such as the European Commission's Marie Curie Fellowship programme are helping them to leave. But scientists who establish themselves elsewhere can sometimes make it back.

**Paul Smaglik**

*Naturejobs* editor



## Contents

### SPECIAL REPORT

Making the move into science  
policy p452

### WWW.NATUREJOBS.COM

Career centre  
Information on the  
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

# Policy-makers in many areas need input from scientists to help them make sound decisions. Virginia Gewin follows the paths to science policy.

For Leah Goldfarb, an invitation to work in science policy was written in the sky. After following the negotiations for the Montreal Protocol — the landmark international agreement to protect the stratospheric ozone layer — she chose to do her PhD work at the University of Colorado, Boulder, on atmospheric chemistry. “I knew I wanted to do science that was relevant to society,” says Goldfarb.

Working with UC Boulder faculty involved in national and international discussions on climate and ozone kindled her desire to work in science policy rather than policy-relevant science. Work with a climate-modelling research group got her to Paris, where she soon found her present position as a science officer for the environment and sustainable development at the International Council for Science (ICSU). As problems such as climate change and desertification become more apparent, many European scientists are taking an interest in science policy. But how do they find work in this area?

Like Goldfarb, many start out doing science for organizations involved in social issues, before their concern leads them towards helping to shape policy. Scientific societies, international organizations, charities, lobby groups, government bodies and non-governmental organizations all need specialists working on science policy.

Others join independent think-tanks, to advise official bodies from the outside. Some even create their own science-policy positions by setting up a

think-tank, or acting as independent advisers.

But before anyone takes the plunge into policy, veterans say they should consider what they wish to do in the long term: whether they want to move back and forth between science and policy, or make the decision to leave one world for the other.

## SHORT TERM

One way to test the water is to take a short-term position. These are often available for scientists with specific technical knowledge and policy interests. This may also offer the opportunity to work abroad, with international organizations such as the United Nations (UN) and the World Bank.

Thomas Rosswall, executive director of ICSU, applauds the increasing number of opportunities for people with experience in policy-relevant, interdisciplinary science. Short-term positions at ICSU come up periodically and offer training in international policy. As the invited ‘voice’ of the scientific community at the World Summit on Sustainable Development meetings in Rio de Janeiro, ICSU enables trainees to get a bird’s-eye view of policy development and become part of the network of international players.

The UN Educational, Scientific and Cultural Organization (UNESCO) recruits scientists for short-term, project-based positions that deal with legislation or public awareness of specific issues. Opportunities vary by country. “Many parliaments do not even have science committees,” says Mustafa El Tayeb, director of UNESCO’s science analysis and policies division.

The European Parliament’s Scientific and Technological Options Assessment office offers a limited number of opportunities for short-term work experience. Its Ramón y Cajal scholarships are available to scientists and engineers, but unpaid research visits can be arranged as well.

Some countries are promoting interactions between scientists and policy-makers at the national level. In the United Kingdom some ten formal fellowships, each lasting three to six months, are funded by scientific societies, research councils and the Parliamentary Office of Science and Technology. The fellows provide briefings on scientific issues to members of parliament (MPs).

Another option for experienced scientists is to be appointed a committee specialist for the House of Commons or the House of Lords, to advise their

Betty Williams MP in the lab at Bangor University as part of this year’s politician–scientist pairing scheme sponsored by the Royal Society.



## Well-established US paths

**Opportunities for scientists interested in policy often follow more direct paths in the United States (see *Naturejobs* 415; 4–5, 2002). A number of scientific organizations, such as the American Chemical Society, support annual short-term positions for scientists and graduate students who want some experience in public policy.**

**In addition, the American Association for the Advancement of Science (AAAS) coordinates the**

**congressional fellowship programme for more than 30 science societies and federal agencies. The National Academy of Sciences also sponsors annual internships for graduate students in both the natural and social sciences.**

**In many instances, the experience is career-changing. Roughly one-third of AAAS participants move into policy-related positions.**

V.G.



For Leah Goldfarb, influencing policy won out over the lab.



specialist committees. These are full-time paid positions lasting up to four years.

The Royal Society has set up a short-term pairing scheme in which a scientist shadows an MP for one or two weeks. In turn, the MP visits the scientist's lab. The hope is that each will then better understand the demands and constraints of their two worlds. Researchers will appreciate the parliamentary timetable, while the MP will become acquainted with the scientific community and its work.

Researcher-led think-tanks have popped up all over Europe to provide policy-makers with scientific advice. The US think-tank RAND has been advising policy-makers through research and analysis since 1948, and has European branches in Cambridge, Berlin and Leiden, the Netherlands. Relative newcomers include the Tyndall Centre for Climate Change Research in Norwich, UK; the Wuppertal Institute for Climate, Environment and Energy in Wuppertal, Germany; and the International Institute for Applied Systems Analysis in Laxenburg, Austria.

These are creating jobs. Mike Hulme, executive director of the Tyndall Centre, hires a few postdocs each year to conduct research. Within the entire organization — a collaboration of nine research institutions and three UK research councils — there can be up to 80 postdocs. Research at Tyndall includes assessing viable options for reducing carbon dioxide emissions and examining society's options for adapting to unavoidable climate change.

#### MIXED TERMS

Opinions differ over the usefulness of scientists leaving science for policy. "Looking at the big picture — you don't want transfers of people from science into policy, you want a temporary shift," says Jonathan Grant, a health-policy specialist at RAND. "You want circulation." Some say that the most useful result of short-term exposure is for scientists to return to the bench with a greater understanding of policy needs.

Those who stay in policy may have to make sacrifices. "There is clearly a lack of structure, and little professional incentive for developing your career," says David Stanners, programme manager for strategic development and international cooperation at the European Environment Agency. But for some, such sacrifices are worthwhile. Goldfarb was happy about leaving the lab, but she passes on a warning she received when she switched from atmospheric chemist to science-policy officer. "Once you leap into science policy it's hard to get back into science," she says. ■

Virginia Gewin is a freelance writer in Corvallis, Oregon.

International Council for Science

♦ [www.icsu.org](http://www.icsu.org)

United Nations Educational, Scientific and Cultural Organization

♦ [www.unesco.org](http://www.unesco.org)

World Health Organization

♦ [www.who.int](http://www.who.int)

United Kingdom Parliamentary Office of Science and Technology

♦ [www.parliament.uk/post/home.htm](http://www.parliament.uk/post/home.htm)

European Parliament Scientific and Technological Options Assessment

♦ [www.europarl.eu.int/stoa/scholars/default\\_en.htm](http://www.europarl.eu.int/stoa/scholars/default_en.htm)

## Take the initiative

**Many short-term opportunities in Europe are under-publicized, so some administrators at scientific organizations encourage those interested to contact them directly for information.**

Reinhard Grunwald, secretary general of the DFG, Germany's main research funding agency, encourages students to take the initiative. "Somebody interested in policy should write to research centres, the

**DFG or the Max Planck Society and tell them what your interests are."**

Mustafa El Tayeb, director of the science analysis and policies division of UNESCO in Paris, agrees. "We want to have our databases updated with interested people," he says. He invites those interested in short-term positions involving legislation and public awareness with UNESCO to contact him directly.

V.G.

♦ [www.unesco.org](http://www.unesco.org)